Classes de dynamiques neuronales et correlations structurées par l'experience dans le cortex visuel.

David Colliaux

Thèse pour l'obtention du grade de

Docteur de l'Ecole Polytechnique

Spécialité : Sciences cognitives

Présentée par

David Colliaux

Sujet

Classes of neuronal dynamics and experience dependent structured correlations in the visual cortex.

Soutenue le 31 Mai 2011 devant la Commission d'examen :

M. Bruno CESSAC	Président du jury
M. René DOURSAT	Rapporteur
M. Khashayar PAKDAMAN	Rapporteur
M. Andrew DAVISON	Examinateur
M. Yves FREGNAC	Co-directeur de thèse
M. Jean PETITOT	Co-directeur de thèse

Préparée à **CREA - UNIC**

Abstract

Neuronal activity is often considered in cognitive neuroscience by the evoked response but most the energy used by the brain is devoted to the sustaining of ongoing dynamics in cortical networks. A combination of classification algorithms (K means, Hierarchical tree, SOM) is used on intracellular recordings of the primary visual cortex of the cat to define classes of neuronal dynamics and to compare it with the activity evoked by a visual stimulus. Those dynamics can be studied with simplified models (FitzHugh Nagumo, hybrid dynamical systems, Wilson Cowan) for which an analysis is presented. Finally, with simulations of networks composed of columns of spiking neurons, we study the ongoing dynamics in a model of the primary visual cortex and their effect on the response evoked by a stimulus. After a learning period during which visual stimuli are presented, waves of depolarization propagate through the network. The study of correlations in this network shows that the ongoing dynamics reflect the functional properties acquired during the learning period.

Keywords: Neuronal networks, dynamical systems, visual cortex.

L'activité neuronale est souvent considérée en neuroscience cognitive par la réponse évoquée mais l'essentiel de l'énergie consommée par le cerveau permet d'entretenir les dynamiques spontanées des réseaux corticaux. L'utilisation combinée d'algorithmes de classification (K means, arbre hirarchique, SOM) sur des enregistrements intracellulaires du cortex visuel primaire du chat nous permet de définir des classes de dynamiques neuronales et de les comparer l'activité évoquée par un stimulus visuel. Ces dynamiques peuvent être étudiées sur des systèmes simplifiés (FitzHugh-Nagumo, systèmes dynamiques hybrides, Wilson-Cowan) dont nous présentons l'analyse. Enfin, par des simulations de réseaux composés de colonnes de neurones, un modèle du cortex visuel primaire nous permet d'étudier les dynamiques spontanées et leur effet sur la réponse à un stimulus. Aprs une période d'apprentissage pendant laquelle des stimuli visuels sont prsentés, des vagues de dépolarisation se propagent dans le réseau. L'étude des corrélations dans ce réseau montre que les dynamiques spontanées reflètent les propriétés fonctionnelles acquises au cours de l'apprentissage.

Mots-clés: Réseaux de neurones, systèmes dynamiques, cortex visuel.

Contents

Introduction

Neuronal activity is often studied through the response of brain networks to a stimulus and in early studies of congnitive neuroscience, ongoing activity was neglected as a disturbance but recent studies highlitghted the potential functional role of these internal dynamics. It is thus important to characterize on going dynamics and to consider their relation to the structure and function of brain networks. We will adress these questions by analyzing and modelling the neuronal dynamics of the cat primary visual cortex.

Advances and diversification in measurement techniques produced a huge amount of heterogeneous data representing brain dynamics. Models of these dynamics also generate a huge data flow which can be analyzed with the same methods as for biological data. A first problem is to classify the dynamics so that a simple request in a database returns recordings of the same class. Thoses classes are also necessary to facilitate the sharing of data by creating a common language and the International Neuroinformatics Coordination Facility aims at defining such standards in neuroscience. Another issue with on going activity is that recording sessions may be long and it is difficult to identify the interesting parts of the sample. An appropriate compression of the signal would then make navigation easier.

The large amount of data available drive modelling and the generativity of neuroscience for producing new mathematical problems is astonishing. Those models range from the detailed reconstruction of brain networks requiring heavy computational ressources to reduced models which captures essential features in an abstract manner and can be analyzed with dynamical systems theory. The Blue Brain project at Lausanne is an example of large scale model including many intrinsic properties and the precise morphology of the neurons and similar projects are currently launched in Europe (Brainscales), USA (SyNAPSE) and China (CABA). In the Brainscales project, those detailed models are combined with mean field analysis to provide a multiscale description.

The analysis and modelling of neuronal dynamics may have several applications, it can be used to predict the effect of drugs or stimulation, it can also be used to implement biologically inspired systems and to design brain machine interfaces.

The organization of the brain may be studied directly by tracing anatomical patterns of connectivity or indirectly through the correlation and coherence analysis of the brain activity and there is now huge efforts to relate these two

measures in the neuroscience community through joint measures and theoretical modelling. The brain is organized into connected clusters of cortical areas forming complex networks [1], neither regular nor random, which can be characterized using methods from graph theory [2]. Evidence for small world properties, constiting in a high clustering index and short average path length, has been found and highly connected hubs in these networks can be identified. These structural properties have functional advantage for efficient communication between areas and explain the complex dynamics observed at a large scale.

The visual system is composed of multiple levels organized hierarchically with each unit receiving feedforward from lower levels, feedback connections from higher levels and lateral connections from units of the same level. The primary visual cortex, also referred to as V1 and corresponding to the cat area 17/18, is the first cortical stage of visual processing. Inputs resulting from visual stimulation at the retina are transmitted through the lateral geniculate nucleus (LGN) in a feedforward fashion. Neurons of the visual system respond preferentially to a particular area of the visual field and the size of this area gets larger and more complex from the retina to the higher levels of the visual system. LGN cells have circular receptive field with high response when a light stimulus hits its center and the surround is dark for On cells, Off cell have the opposite property of being highly activated by dark center and light surround. The simple cells of the primary visual cortex are selective to a specific orientation with the shape of a 2 dimensional Gabor wavelet with alternating On and Off bands. Such selectivity is characterized by an orientation tuning curve representing the firing rate response of the cell as a function of the orientation of the presented bar or Gabor patch. It has been shown that orientation tuning is widespread among mamals and similar functional properties are found across species despite major differences in the local connectivity [3]. For primates, having a richer visual environment than rodents and cats, receptive fields are diverse and may contain more alternating On and Off bands than for other animals [3]. Cells in higher areas like the inferotemporal cortex may be selective to very specific features [4] and place cells in the hippocampus is active when the animal is located in a small area of the environnement.

In the primary visual cortex of the cat, cells with receptive field corresponding to close points of the retina fall close together so that a 2 dimensional retinotopic map over the surface of V1 represents the visual space. This topographic organization can also be found for other sensory modalities like in the auditory where sounds are represented by a tonotopic map with cells coding for close frequencies located close apart. On the surface of the primary visual cortex, the preferred orientation of cells vary smoothly turning around special points of the map referred to as pinwheel singularities. Other features like the spatial frequency and the direction of motion of the stimulus are coded in V1. Lateral connections tend to favor cells with similar orientation properties and the combination of the pinwheel map with these connection pattern constitute a functional architecture implementing low level visual processing including feature integration and pattern completion [**?**].

In the vertical direction, cells are densely connected and have similar receptive field properties both in position and orientation, providing robustness of the code through redundancy. These functional columns are ubiquitous in the brain from sensory areas to higher levels like the prefrontal cortex involved in working memory and cognitive control. Thus functional columns may be considered as mesoscale computational units supporting distributed operations of various sorts. Actually, columns have different properties depending on the animal species and the cortical areas considered [5] making the ¡¡tabula rasa¿¿ view of a uniform brain a crude assumption. Moreover, there are six layers in the depth of the brain forming local circuits. Feedforward connections from the LGN, for example, terminate onto layer IV. In abstract models, these specific connections may be replaced by random connections with uniform probability when the layered structure is not relevant. The graph of neurons in the primary visual cortex is thus topographic, modular and random.

Brain dynamics can be described at multiple temporal and spatial scales and it is important to identify its relevant features associated with the coding of sensory information, the memory maintenance and the behavioral state. At the single cell level, spikes are thought to convey information either by the firing rate or their temporal relations to spikes of other cells. The firing rate of a simple cell in V1 is thus at its maximum when the stimulus is oriented along the preferred orientation. With such a rate coding, the timing of spikes is not taken into account but some experiments showing fine precision suggest evidence for a temporal coding [6] [7].

At the population level, the code is independent or correlation based depending on whether temporal relationships between spike trains are taken into account. Synchronous firing resulting from oscillatory modulation of the membrane at the gamma frequency (40Hz) in the primary visual cortex of the cat was found to implement the binding of parts of an object into a whole in 1989 [8]. Neuronal oscillations were then described in other areas and involved in many cognitive processes like attention or memory [9]. Place cells in the hippocampus are transiently binded into a synchronous cell assembly oscillating in the theta band (8Hz) when a memorized place is recalled. Complex oscillatory activity generated by specific circuits called central pattern generators is associated with rythmic behavior like respiration, heartbeat, locomotion or chewing.

Synchronized transitions between a quiescent down state where the membrane potential is hyperpolarized and a depolarized up state where the neurons spikes are related to working memory in the prefrontal cortex with persistent up state representing an object hold in memory [10] [11]. Several dynamics are found in the up state, firing may be plateau-like, oscillatory or ramping [12]. Slow transitions between up and down states in the thalamus and the cortex are also related to the slow wave observed during slow wave sleep or anesthesia [13]. Low dimensional chaotic dynamics are often considered to reflect pathological states like epilepsy but recent findings suggest that it could provide flexible representation used for memory [14] [15] [16] and for example, in the antenna lobe of the fruit fly, odors have been shown to be coded as transient trajectories in an abstract low dimensional phase space [17].

The functional dynamics described above are often recorded at the population level. The firing rate code is thus investigated through the detection of activated regions of the brain with functional magnetic resonance imaging (fMRI) when the subject is involved in a cognitive task. Transient synchrony is also observed at large scale with electroencephalogram (EEG) and it is correlated with binding of a coherent percept, attention or working memory with multiple frequency bands interacting [18]. Spatiotemporal patterns relevant for perception or memory may also be recorded with optical imaging using voltage sensitive dye or with arrays of electrodes. Visual stimulation thus triggers traveling waves in the visual cortex [19] and information processing in sensory areas also generates specific phase patterns of oscillatory activity [20].

For a long time, the brain at rest has been considered as a passive medium in cognitive science and most of the studies focused on evoked activity related to a controlled task and cancelled the intrinsic dynamics through averaging. The observation that most of the energy used by the brain is devoted to the on going activity raised new questions about the structure and the function of default networks having coherent activity when the subject is resting which decreases when the subject is involved in a task [21]. The internal dynamics help the brain to face computational problems despite the poverty of stimulus, for example only a small part of the information impinging the retina reaches the primary visual cortex. The intrinsic activity may implement priors reflecting knowledge about the statistics of the visual world thus supporting theories about vision as an inference process [22]. Default networks are also considered to support internal thoughts and a representation of the self. Intrisic activity is also interesting to study as it explains variability in behavior and in evoked responses.

The ongoing activity defining a global brain state depends on the state of consciousness that is on whether the subject is sleeping, anesthetized or awake. Sleep can be decomposed into cycles of REM and NREM[1] phases. The NREM sleep is composed of three stages having specific dynamic features. Stage 1 is characterized by theta waves (around 7Hz) and in stage 2 spindle waves (around 10 Hz) and K complexes[2] are found, predominantly in frontal areas. During stage 3, also called slow-wave sleep, a slow oscillation in the delta band (around 1 Hz) propagates along the brain correlated with up and down states at the cellular level. The REM sleep is a stage in which most of dreaming occurs and the associated dynamics are similar to those observed in the awake state that is irregular at the cellular with a nearly Gaussian distribution of the membrane potential and with no slow oscillation at the global level measured by EEG [13]. Both REM and NREM dynamics are involved in memory consolidation whereas dynamics of the awake state are involved in information processing and mental representations. The default network has been recently identified via the BOLD signal of fMRI but the corresponding neuronal dynamics remain an active subject of research [23]. Anesthesia have different effects on the on going dynamics depending on the drugs used. Xylazine-ketamine thus have a drastic

[1] Rapid eye movements (REM) and non rapid eye movements (NREM).
[2] Short negative peak followed by a slower positive peak.

effect producing synchronized up and down states close to those observed during slow wave sleep and alfatesin or urethane producing asynchronous and irregular activity close to those observed during the awake state [24]. In the primary visual cortex, it has been found that the ongoing dynamics itinerates among states corresponding to the activity evoked by a specific orientation so that the spontaneous dynamics reflect features encoded in the map [25].

In order to model the brain dynamics listed above, several approaches can be considered but they all have to include a dynamical system defining the units of the network and a functional architecture to connect these units. With a bottom-up approach, the aim is to build biologically realistic networks including as much details as we know from studies in neuroanatomy and neurophysiology. This bottom up approach is computationally demanding and requires a lot of data but improvements in these domains will soon make it possible to build artificial neuronal networks with the size of the human brain (around 10^{11} neurons). The phase space and the parameter space of these models is so huge that it is difficult to control and to analyze its dynamics. With a top down approach, the simplest phenomenological models are considered to reproduce some features of the observed dynamics like synchronization. A model can be made more realistic by considering a more complex model for the neuron, adding some dynamics on the connection weights to model synaptic plasticity or using a more complex architecture. A model with a bottom up approach can be reduced to a simpler form by considering equivalent or average equations in a phase space with lower dimension or relying on hypotheses like considering cells to have the same parameters in a population. Dynamical systems can be used to describe networks of very few units or it should be combined with statistical physics to study very large systems through their mean field description.

Models taking a spatial extension into account for the neurons are computationally heavy because they require the integration of partial differential equations in a 3 dimensional space and a common approximation is to consider space clamped models, also called point neurons, where the dynamics is only considered at the soma without modeling the propagation of the spike along the axon. Detailed model of the dynamics at the soma, like the Hodgkin-Huxley, include voltage gated channels and a leak current. The essential features of the dynamics in a neuron model like spiking and bursting can be studied using bifurcation theory and this analysis leads to reduced models [26]. A common method is to build an oscillator based on the computation of the phase reponse curve and this reduced system has the same synchronization properties as the original system. Another simplification is to forget the precise wave form of the action potential and to replace it with an instantaneous reset and a refractory period resulting in hybrid dynamical systems which combine a continuous and a discrete part. In the most drastic approximation, the neuron is a two states system, it is spiking or quiescent depending on whether its total input is more or less than a threshold.

In data-driven models, quantitative maps from studies in neuroanatomy are used to build the network [27] whereas with a top down approach generic structures are considered, like a homogeneous population of randomly connected neu-

rons either sparsely or densely. Repeated patterns of connectivity can be used as building blocks for a modular structure, two connected populations, one excitatory and one inbhitory, are thus a generic model for the column and these columns can be gathered to form hypercolumns or maps [28]. Some architectures have been designed specifically for information processing. For Hopfield networks, composed of binary neurons connected symetrically, an energy function can be defined [29]. The minima of this energy function are fixed point attractors for the dynamics which can be used for building associative memory. As an input brings the system in the basin of attraction of fixed point, the dynamis of the network converges toward this fixed point, performing pattern completion and providing a way to store represention coded in the connectivity matrix. Synfire chains were designed with multiple layers of neurons connected in a feedforward fashion to test the ability of neuronal networks to transmit synchronous volleys of spikes. If there are enough spikes in the input volley and if the jitter in their timing is small enough, such architecture can propagate synchronous activity in a stable fashion even with a background activity [30]. A generalization for synfire chains are the polychronous groups which are repeated patterns of spikes, including possible delays [31]. The liquid state machine is also a neuronal architecture designed to perform computational task, it is composed of three parts: time varying inputs are connected to a recurrent neuronal networks and a readout combining neurons from the recurrent network. The central recurrent network maps the inputs to a very high dimensional space and connections from this reservoir of dynamics to the readout can be learned to perform a specific task like discrimination between two possible inputs [32].

Although some transient coherent oscillations may be observed, the ongoing activity in cortical networks associated to the awake state is highly irregular and asynchronous. Sources of randomness in neuronal dynamics may be attributed to intrinsic noise in ionic channels or stochasticity in synaptic transmission but also to a chaotic regime that has been found in sparsely connected networks of binary neurons when inhibitory inputs balance precisely the excitatory inputs [33]. This self-sustained asynchronous activity was later obtained in networks of integrate and fire neurons with conductance base synapses [34] or in dense networks [35]. Up and down states, as found during slow wave sleep, have also been modeled in neuronal networks [36] [37], collective transitions between up and down resulting from the recurrency of the network and intrinsic properties of the neuron like a non linear current inducing bistability of the membrane potential or spike frequency adaptation.

In this thesis, we will propose a method of analysis for the ongoing activity in the primary visual cortex, we then study reduced models for neuronal dynamics and in the final part such reduced models are used in a model of the primary visual cortex. We show how an asynchronous irregular state with a Gaussian distribution of the membrane potential at the single cell level can be switched to state with experience dependent structured correlations with a bimodal distribution for the membrane potential.

In the first chapter, we present a collection of tools to analyze temporal signals in neuroscience with a special interest for spike trains, membrane po-

tentials and EEG. Spikes are an essential feature of neuronal dynamics as they are related to information processing and spike trains are characterized by their mean frequency and their regularity. The naked membrane potential, that is the membrane potential from which spikes have been removed, is also investigated trough the moments of its distribution, the comparaison of its distribution with the closest Gaussian distribution and measures of the bimodality of its distribution. Spectral properties of analogous signals are described and a kind of time frequency analysis is introduced via wavelet transform. Based on the compression of the time frequency matrix, we propose an original method for building the graph of transitions among microstates of the EEG. Signals are compared using a distance or a correlation coefficient and specific distances have been recently designed for spike trains. Information theoretic measures are promising for the exploration of multi-channels data as can be recorded using an electrode array but we point at the technical difficulties in estimating the entropies and the heavy computations they require. Finally, attractor reconstruction is a classical method in EEG which can also be applied to the analysis of the membrane potential. When a set of parameters is measured for each signal, unsupervised learning methods are applied to find a structure in the data set and we present three possible methods: K-means, hierarchical tree and self organized maps. Each method having advantages and pitfalls, we build a similarity measure combining results from several classification methods and which we call metasimilarity and we also propose a method to compare partitions resulting from different methods. Classification methods and the metasimilarity measure are first tested on a set of random samples generated from a combination of Gaussian distribitutions. We then apply these methods to classify 150 electrophysiological recordings of the ongoing activity in the primary visual cortex of anesthetized cats obtained in the last 10 years at the UNIC. For each recording, 25 parameters were selected and classification methods were applied to the obtained data set after normalization and dimension reduction. With K-means clustering, we obtain an optimal partition into 6 clusters which define classes of neuronal dynamics. For most of the cells, both ongoing and evoked activity are available so that we can check how visual stimulation change the dynamics and we conclude that it reduces the accessible dynamics. Nazied Huguet and Cyril Monier contributed to this part by selecting and computing relevant parameters for each cell.

In the second chapter, the computational units of the brain, the neuron and the cortical column, are presented with methods to analyze their dynamics. We give a short introduction to dynamical systems, either deterministic or stochastic, including bifurcation theory and the Fokker-Planck equation. A first part is dedicated to neuron models, the Hodgkin-Huxley model is commonly used to describe the evolution of the membrane potential but its 4 dimensional phase space and its nonlinearities make it quite complicated to study. It can be reduced to a two dimensional system, the FitzHugh-Nagumo model, or hybrid dynamical systems of various types also called integrate-and-fire models. We make a complete analysis of local bifurcations of codimensions 1 (fold and Hopf) and 2 (Bogdanov-Takens and Bautin) in the FitzHugh-Nagumo model and we

study the effect of noisy stimulation at a codimension 3 bifurcation point, the organizing center of the system. Classical results about the integrate-and-fire model under deterministic and stochastic forcing are recalled and a recent variation including an exponential non-linearity and a slow adaptation variable is presented. In the last part of the chapter, we focus on models of the cortical column and small and generic networks of coupled columns. A column consists of two homogeneous populations, one excitatory and one inhibitory, and bifurcations of its mean field equations are well known since the work of Wilson and Cowan. When the networks is only sparsely connected, an asynchronous irregular state can be stable and a column of integrate and fire neuron may also have synchronous irregular or synchronous regular dynamics depending on its input and its balance between inhibitory maximal conductance and excitatory maximal conductance. When the network have oscillatory activity, it can be reduced to a phase oscillator, the synchronization properties of few coupled columns are captured in such phase dynamics and in the limit of a large population of globally coupled, it is the Kuramoto model. A ring of oscillators with long range coupling exhibit dynamics in which a part of the network is phase locked and the other part is desynchronized. We shortly discuss a model combining a voltage equation and an oscillatory current so that the system has a quiescent down state and an oscillatory up state. We studied this model at the DEI laboratory (Riken,Tokyo) under the supervision of Yoko Yamaguchi and in collaboration with Colin Molter. In the publication in appendix, we study the synchronization of two coupled units and we find a small window of perfect phase synchrony followed by chaotic behavior and a network of such units. These units can also be used to generate working memory dynamics in a network embedding cell assemblies and we show that the ongoing dynamics itinerates among the stored memories.

The third chapter is devoted to rate models of the formation of V1 and large scale networks of spiking neurons modeling a small patch of the primary visual cortex. We present four approaches to the formation of V1 some based on dynamics in a feature space and some based on the full dynamics of the units activity and connection weights. The cognitive approach is based on the modular theory of the brain, the brain is composed of many building blocks at a low level which can be grouped together to form a functional unit at a higher level. In the Hubel and Wiesel model for orientation selectivity in V1 for example, aligned receptive fields in the LGN are grouped to form a simple cell receptive field. The functional view defines a goal to the formation of V1 which is an optimal mapping from the surface of the cortex to the feature space where inputs are represented and variational formulation gives the dynamics to solve this problem. The elastic net is an example of such an algorithmic approach. With a physical approach, orientation maps and pattern of ocular dominance results from universal mechanisms of pattern formation and phenomenological models can give accurate prediction for the wavelength of ocular dominance bands and the density of pinwheels. Finally, in a biological approach, a learning mechanism, based for example on the Hebb rule, gives some dynamics on the connection and the functional architecture gets formed as visual inputs are pre-

sented to the network. Large scale networks of spiking neurons are implemented using the pyNN interface to the NEST simulator to study the fine scale dynamics of neurons in models of V1 with a modular structure. Pierre Yger contributed to the implementation of these models. In the first model inspired by the Hubel and Wiesel model, a regular grid of 11×11 columns, each containing 50 neurons, is used for V1 and a grid of 11×11 is a lumped model for the Retina-LGN inputs to V1, feed-forward connections from the inputs are hard wired to form simple cell receptive fields. We study the possible regimes of ongoing activity depending on the maximal conductances in the network and the response to an oriented stimulus, either static or rotating, depending on the ongoing state. In the second model, we take a larger grid (30×30) and long range isotropic connection kernels for lateral connections. We describe the ongoing dynamics of the network and we find that for a range of the maximal conductances some static pattern emerges similar to the classical hallucination patterns from rate models. In a region of the parameter space, a focal stimulation of the network gives rise to a traveling bump, which is also a classical solution for rate models. The connectivity of the third model is obtained after learning under presentation of oriented stimuli in a rate model, the connection weights in the macroscopic rate model are considered as the probability of connections between neurons of the corresponding columns in the network of spiking neurons. When the connectivity of the second model is used to initialize learning, an orientation map emerges and the dynamics in the network of spiking neurons can be compared before and after learning. Between the asynchronous irregular state and the uniform saturated state, we find a region of the parameter space where neurons in a column have collective transitions between up and down states and we show that, in this regime, spike correlations depends both on the distance and the difference of preferred orientation between two columns. We thus show how the primary visual cortex may be switched from a state where neurons fire in an asynchronous and irregular fashion to a state where experience dependent structured correlations are propagated in the network resulting in up and down states at the column and neuron levels.

I would like to express my sincere gratitude to Jean Petitot from the CREA and Yves Fregnac from the UNIC, who supervised this PhD. I got my first interest for neurogeometry in the teachings of Jean Petitot and it has been a very exciting adventure to confront theoretical apects with experimental data and computer simulations with Yves Fregnac's team at the UNIC. I would also like to thank them and Paul Bourgine from the CREA for their kind support.

I benefited from a very nice and stimulating research environment so I warmly thank researchers and the administration staff of CREA-ISC and UNIC for that. I would especially like to mention people I have been working with closely: Nazied Huguet and Cyril Monier from the UNIC who contributed to the analysis of neuronal dynamics, Gilles Wainrib from the CREA with who I had stimulating discussions about dynamical systems applied to neuroscience and Pierre Yger from the UNIC who made me like Python and contributed to the implementation of large scale networks. I also thank Jonathan Touboul for fruitful discussions about bifurcation theory and Matlab and Laurent Perrinet

for his comments on the last version of the manuscript.

I had the opportunity to join the RIKEN BSI summer school in Tokyo at the begining of my PhD as my first research experience. I would like to thank researchers of the DEI laboratory, especially Yoko Yamaguchi, who supervised the internship, and Colin Molter for fruitful discussions about cognitive neurodynamics.

I am also glad to thank friends without who coffee wouldn't taste so good: Antoine, Ben, Bertrand, Geoffroy, Nicolas, Raphaël, Valérie et Victor.

Finally, I would like to thank my family for their constant support and Misa for her kind attention.

Chapter 1

Phenomenology of cortical dynamics: All you need to know about your data

1.1 Introduction

In neuroscience, experimentalists are confronted with a huge amount of data of very different nature. At the same time, given a good model, it is easy to reproduce realistic dynamics mimicking those signals. For example, it is possible to produce the output of a neuron given its input with great fidelity. The simulations obtained by computer scientists also generates a huge amount of data and the resulting signals are very close to those recorded in biology. The similarity of these artificial and natural data suggest that the same methods of analysis should be used. We present in this chapter a collection of tools and techniques which can be used to analyze and classify signals in biological and computational neurosciences.

The first part is an introduction to the common representations of the brain activity that are the spike train, the membrane potential of a neuron and the EEG. The dynamics at the single cell level is characterized by static properties related to the distribution of the membrane potential, spectral properties and firing properties. We also describe more sophisticated measures like based on information theory to manage signals from multiple channels and attractor reconstruction which found applications in the analysis of macroscopic signals. A method based on time frequency analysis is proposed to compress long recordings into a sequence of states and a graph representation of these states and their transitions is provided. In the second part, three classification algorithms are described: K-means, hierarchical tree and self-organized maps and we propose some methods to compare and combine them, thus avoiding the pitfalls inherent to each algorithm. The analysis techniques described in the first two parts are

applied, in the third part, to single cell recordings of the ongoing activity in the primary visual cortex of anesthetized cats. Each data sample is represented by 25 parameters and a clusterization in this parameter space gives an optimal partition into 6 clusters. Under visual stimulation, the same cells gathers in the main cluster so that we find more accessible dynamics in ongoing activity than in the evoked activity.

Classes of neuronal dynamics are classically defined by the response of a neuron to a stereotyped electrical stimulation, this study aims at the definition of new classes based on the ongoing and visually evoked activity.

1.2 Temporal signals in neuroscience

The nervous system is considered from Galien to Descartes by an hydraulic analogy with a nervous fluid flowing in the pipes of the nervous system. The electrical nature of the flow in the nervous system was first demonstrated by Luigi Galvani in Bologna at the end of the 18^{th} century. He reported in 1791 that an electrical stimulation of a nerve fiber of a frog could generate a muscle contraction in its leg and, in 1797, he reported that the same contraction could be obtained by pulling to nerve fibers together suggesting the first evidence for animal electricity production. During the 19^{th} century, galvanometers became more and more precise to detect electrical signals and German physiologists, like Emil du Bois-Reymond, could characterize the nervous signals as constituted of short depolarizing events. At the end of the 19^{th}, the physico-chemical mechanisms responsible for this signal were better understood with for example the electro-chemical law giving the potential difference resulting from ion concentrations inside and outside the cell, now known as Nernst potential. With the giant squid axon, Hodgkin and Huxley found, in the 30's, a nerve fiber thick enough to record its activity with a microelectrode clamped to the neuron and this led to their seminal work of the 50's were they described precisely the action potential and proposed the model for its generation. This led to modern electrophysiology were the membrane potential with spiking activity and synaptic events is now recorded in many animal preparation. Using a thicker electrode, the population activity can be recorded and depending on the impedance of the electrode and the filtering of the signal, the recorded activity can reflect the mean depolarization in the dendritic tree or the spiking activity of a set of neurons. By using matrices of such electrodes (MEA), few hundreds of neurons can be recorded at the same time. The Electroencephalogram (EEG) is also a macroscopic signal measuring the spatially averaged activity over a large population of neurons. The whole brain activity can be mapped through an electrode array of 64 or 128 electrodes. The rhythms found in this signal are of special interest for cognitive neuroscience. It can used for assessing the level of consciousness of a subject, to detect precursors of an epilepsy crisis and it also have specific patterns depending on the task the subject is doing. Magnetoencephalogram (MEG) complements EEG by measuring the magnetic field produced by currents running tangentially to the surface of the skull. The obtained signal

is easier to localize and less affected by the skull but the measurement must be done in an environment free of magnetic perturbation thus requiring a heavy equipment. More recent techniques to record neuronal activity rely on optical methods. Through calcium imaging the propagation of an action potential can be tracked with fine temporal and spatial resolution. Macroscopic signals obtained from intrinsic optical imaging (IOS) or after the application of a fluorescent dye sensitive to the voltage (VSD) gives a coarse grained picture of the nervous activity in cortical tissues.

In order to analyze the ongoing dynamics in the primary visual cortex of the cat, we will focus on the membrane potential and the intracranial EEG. Those signals are related since EEG signal is an spatial average of the synaptic inputs and collective variations of the membrane potential are correlated with the EEG variations.

1.2.1 Analysis of a spike train

Spikes extraction A temporal trace of the membrane potential V_m recorded at the soma of a cell contains spikes [1] which are short and rare events easily detectable by a human as shown in fig 1.1 and it would bias any processing of the membrane potential. The extraction of these spikes is thus necessary for a simpler description of the membrane potential and a compact representation of the information contained in the spikes.

The spike time is defined as a maximum in second derivative of the membrane potential which correspond to an explosion of the curvature in the trace when the spike is initiated. This maximum is one order of magnitude higher than spurious maxima due to fluctuations in the membrane potential, so that it is easy to detect by requiring to be at least 3 times higher than the standard deviation.

Near the spike time, the shape of the spike can be approximated by a quadratic curve,$V_m(t) = V_m(t_i) + \kappa t^2$ with κ the curvature, or an exponential function, $V_m(t) = V_m(t_i) + e^{t/\Delta}$. An approximation of the spike time precision can be obtained from the curvature,see [38] and [39]:

$$\delta t = \sqrt{\frac{< \delta V >}{< \kappa >}}$$

for $t_i < t < t_s$ where t_s is the time at which V_m reaches the top of the spike and with averages taken over all spikes.

The value of the membrane potential when the spike is initiated is the spiking threshold and the time it takes for the membrane potential to terminate, that is to cross this threshold from top to down, is the spike duration. Spike removal is achieved by interpolating the membrane potential trace between spike initiation and spike termination. In fig 1.1, the interpolation is linear but smoother traces could be obtained by using splines. An efficient way to remove all spikes on a

[1] The mechanism responsible for the generation of those spikes will be detailed in Chapter 2.

membrane potential trace is to calculate the average wave form of the spike and to estimate the spiking threshold and the spike duration on this average spike. The same threshold and the same duration is then used for all spikes in the trace.

Spiking activity. After the spikes have been removed, the spike train and the spike-stripped subthreshold membrane potential (which will be referred as membrane potential for simplicity in the following) can be analysed separately. General methods for the analysis of spiking activity can be found in [40], [41] and more sophisticated methods are described in [42], [43].

The spike train is a vector of spike timings, $\mathbf{t} = (t_i)_{1<i<n}$ of size the number of spikes detected. Actually, knowing whether the absolute value of those times is of special interest is still an open issue but the time between two spike occurrences gives an indication of the level of activity of the neuron. The interspikes interval, $ISI_i = t_i - t_{i-1}$, is used to define the firing instantaneous frequency of the neuron $f_i = \frac{1}{ISI_i}$. The firing rate can be obtained by averaging the spike count over a time window of width τ, $r_\tau(t) = \frac{1}{\tau}\int_t^{t+\tau} \rho(t)dt$ where $\rho(t) = \sum_{1<i<n} \delta(t - t_i)dt$ is the spike train function [2]. When this quantity is averaged over all the time of the recording or on a time window larger than the spike duration, it is called the mean firing rate and when averaged over many neurons it is called the population firing rate. The firing frequency of a neuron depends highly on its cellular type and on the brain area where it is located. In visual cortex, cells fire with an average firing rate around 1 Hz in barrel cortex [3], 5 Hz in the primary visual cortex and 15 Hz for spontaneous activity in higher level areas like motor cortex or prefrontal cortex with up to 80 Hz when it is activated. During a spike, the membrane is insensitive to incoming current so that even when strongly stimulated in artificial conditions, the firing frequency of a neuron is limited at 1000 Hz due to this refractory period of few milliseconds.

Spike trains are digital signals that is series of 0 and 1 and an analog representation of the spike train s is obtained after convolution of a kernel f with the spike train function $s(t) = \sum_{1<i<n} f(t - t_i)$. The commonly used kernels are the exponential kernel, $f_{exp}(t) = H(t)e^{-\frac{t}{\tau}}$, H being the Heaviside function, and the alpha kernel $f_{alpha}(t) = te^{-\frac{t}{\tau}}$. This analog signal provides a realistic approximation of the input current or conductance corresponding to this spike train and, as will be shown in the part "Metrics and measures", it is also used for building spike train metrics.

Spiking regularity The ISI distribution is also useful to quantify the regularity of the spiking activity of a neuron by the coefficient of variation of interspikes intervals $CV = \frac{<ISI-<ISI>>^2}{<ISI>}$ [4]. For a perfectly regular spiking

[2] The Dirac function $\delta(t - t_i)$ is 1 when $t = t_i$ and 0 otherwise.

[3] The barrel cortex is the somatosensory receiving inputs from vibrissae of the rat or mouse

[4] It is thus the ratio $\frac{\text{VarianceofISIs}}{\text{MeanofISIs}}$.

neuron, all ISIs are the same and the CV is 0. For neurons having $CV = 1$, the variance of interspikes intervals is equal to its average. The simplest stochastic process generating spike train with this property is the Poisson process, where ISIs are independents, and it is a commonly used to model irregular trains of events (see Chapter 2), the ISI distribution of a Poisson process follows a Gaussian law. Many cells in the brain fires in a Poissonian fashion, $CV \approx 1$ in the spontaneous regime, but a closer look at the ISIs distribution shows that it is better described with a gamma law [5] than a Gaussian law. A sub-Poissonian ISIs distribution, $CV < 1$, is characteristic of cells having a more regular firing than if its spike train was generated by a poisson process. A supra-Poissonian ISIs distribution, $CV > 1$, is characteristic of cells that tend to fire with bursts of spikes and is found in evoked activity. The slope of the decay in the ISI distribution may also be an important parameter in cells with low frequency spiking because it reflects how rare events occur which is not taken into account in the previously described parameters.

1.2.2 Analysis of a membrane potential trace

Spikes are a major feature of neuronal dynamics but the subliminar activity, that is fluctuations of the membrane potential under threshold, is also very informative. The membrane potential is a very complex signal reflecting the activity of the network in which it is embedded. Bistability of the membrane potential is found in multiple areas of the nervous system. It sometimes result from intrinsic mechanisms like in the Purkinje cells of the cerebellar cortex [44] where it may support information processing or it may collective and rely on network mechanisms, like in the prefrontal cortex where columns have persistent up state during the storage of an object in the working memory. During slow wave sleep those transitions are correlated with EEG variations. The presence of several levels of activity, like an up activated and a down desactivated state, indicates multistability of the network and transient oscillations are a sign of coordinated spiking in the population. The analysis should then be led carefully to detect such events.

Static properties As will be seen in section 4, much of the information about a cell is hidden in its membrane potential distribution. The simplest way to characterize it is to calculate its successive moments of order k relative to the mean μ_{Vm}, $\mu_k = E((V_m - \mu_{V_m})^k$. The Gaussian, used as a reference to compare probability distribution functions, has a finite second order moment and null moments of higher order. It is defined by

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

[5] The gamma law is a two parameters (k, θ) probability distribution function defined as follows:

$$f(x, k, \theta) = x^{k-1} \frac{e^{-x/\theta}}{\theta^k \Gamma(k)}$$

with $x, k, \theta > 0$ and Γ the gamma Euler function.

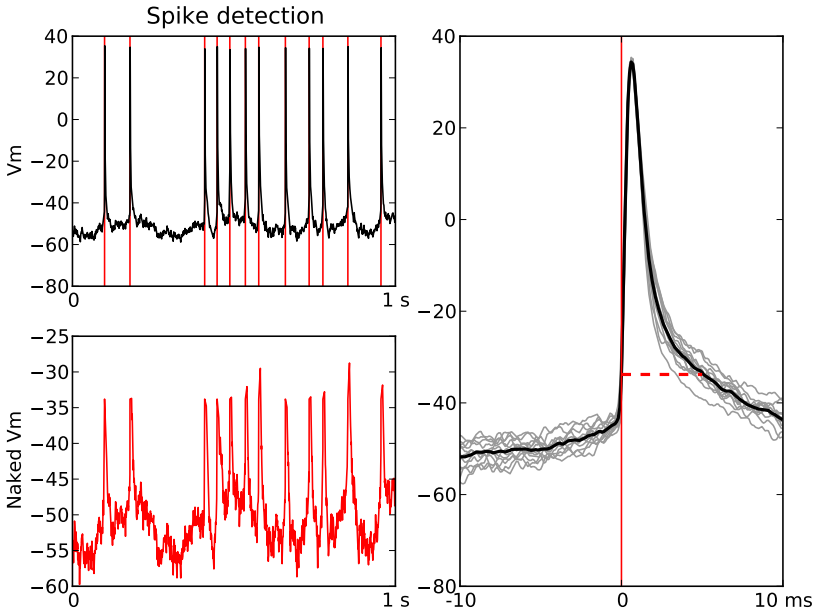


Figure 1.1: **Spike extraction** - (Left-top) Temporal trace of the membrane potential with spike times. (Left-bottom) Trace of the membrane potential after the spikes have been removed. (Right) Average spike for the estimation of the spiking threshold and spike duration.

.

The mean of the membrane potential can be very different from one experiment to another because it depends on many parameters of the experimental preparation. It is usually between -80 and -50 mV [6]. The standard deviation, $\sigma_{V_m} = \sqrt{\mu_2}$, reflects the level of activity in the network. It often depends on the mean μ_{V_m}, there are less fluctuations when a cell is close to threshold than when it is depolarized. The mean and the standard deviation of the distribution are sufficient to fit a Gaussian distribution and the coefficient of regression measures the goodness of the fit. The *skewness*, $\gamma_1 = \frac{\mu_3}{\sigma^3}$ reflects the symmetry of deviations from the mean, it is 0 for a Gaussian distribution. A positive skewness indicates the presence of micro up states as in excitation driven cells and a negative skewness indicates the presence of micro down states as in inhibition driven cells. The symmetry of the distribution can also be checked by using the fitted Gaussian law as a reference and calculating the following coefficients:

$$S_1 = 3\frac{m - \mu_{V_m}}{\sigma}$$

and

$$S_2 = 3\frac{m - med_{V_m}}{\sigma}$$

with m, σ the mean and standard deviation of the Gaussian function and med_{V_m} the median of the empirical distribution. The *kurtosis*, $\beta_2 = \frac{\mu_4}{\sigma^4} - 3$, reflects the sparseness of deviations from the mean, it is 3 for a Gaussian distribution. Distributions with kurtosis greater than 3 are flat and correspond to traces with small and fast fluctuations as would be characteristic of a cell embedded in a very active asynchronous network. Distributions with a kurtosis less than 3 are sharp and corresponds to cells with slow and large deviations from the mean as would be characteristic of a network with low but synchronous activity.

A distribution F is unimodal if there exists a mode m such that F is convex on $[-\infty, m[$ and F is concave on $]m, \infty[$. If the distribution is multimodal that is if it contains more than one peak, the Gaussian distribution is not a good approximation anymore and the distribution can be fitted with a sum of two or more Gaussian laws. For bimodal, the upper peak defines an up state and the lower peak defines a down state. The minimum of the distribution between those two peaks is the threshold separating the up domain from the down domain. Several parameters can be used to characterize deviations from unimodality of a distribution. The distance between an empirical distribution and a test distribution is $\rho(F, G) = sup_x|F(x) - G(x)|$ and the dip of F is $d = inf\rho(F, \mathcal{U})$ where $\mathcal{U}$ is the set of unimodal distributions. A practical way to perform this calculation is described in [45]. The separability is defined from the fit of a sum of two Gaussian functions as

$$Sep = \frac{m_1 - m_2}{2(\sigma_1 + \sigma_2)}$$

[6] The membrane potential is bounded from below by the potassium inversion potential and from above at 0 mV by the Na inversion potential.

with m_1, m_2 the means of the two Gaussian functions ($m_1 > m_2$) and σ_1, σ_2 their standard deviations. The contrast between the two distributions, also called the discretness, is defined as follows from the two Gaussian functions resulting from the fit:

$$Discr = 100 * \sum_i \frac{|G_1(x_i) - G_2(x_i)|}{G_1(x_i) + G_2(x_i)}$$

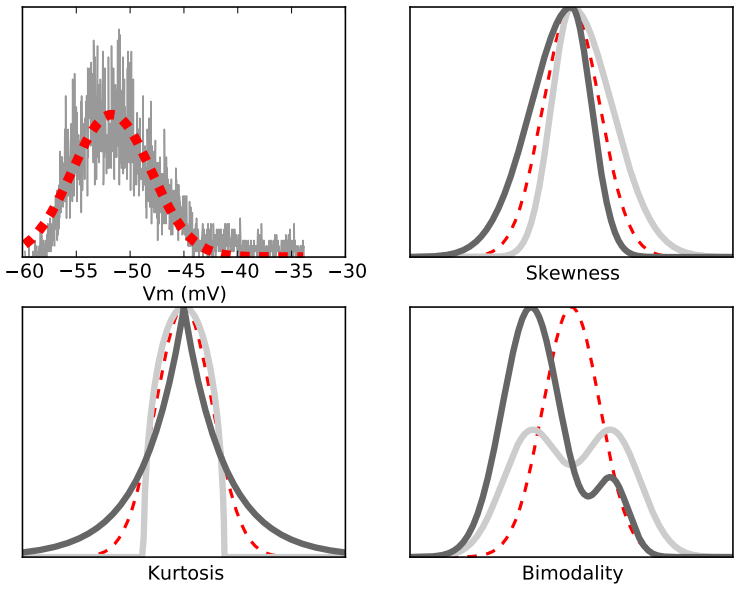


Figure 1.2: **Static properties of V_m** - (Top-left) Gaussian fit for the V_m distribution of cell X. (Top-right) Examples of distributions with positive (dark) and negative (light) skewness. (Bottom-left) Examples of distributions with kurtosis greater than 3 (dark) and less than 3 (light). (Bottom-right) Examples of asymmetric (dark) and symmetric (light) bimodal distributions.

Spectral properties

Autocorrelation Oscillatory behavior of the membrane potential is not detected by the analysis of distribution and transitions between up and down states. There are several possibilities regarding the origin of these oscillations. The whole network can be oscillating in a robust manner at low frequency, this is the case when the brain is in deep sleep, also called slow wave sleep, or when it

is in a pathological state like epilepsy. Transient oscillations at higher frequency can also be seen, and are often considered as the propagation of coherent activity among the cell assemblies in which the neuron is embedded. A simple way to detect oscillations in a signal s is to calculate its autocorrelation,

$$R_s(\tau) = \frac{1}{T}\int_0^T s(t)s(t-\tau)dt.$$

A first time constant is given by the extinction rate, τ_e, which can be captured by fitting an exponential function, $e^{-\frac{t}{\tau_e}}$. In the case of cell X, the autocorrelation decreases in a linear fashion. It is still possible to see a slight oscillatory deviation from the linear behavior at $\tau \approx 50ms$, which is close to the average ISI of the spike train.

Power spectral density (PSD) To get more information about the frequency content of the membrane potential fluctuations, it is interesting to calculate the power spectral density and this is done by using the Fourier transform of the signal. The Fourier transform of a signal is

$$\hat{s}(\omega) = \frac{1}{T}\int_T s(t)e^{i\omega t}dt$$

and the PSD is then $S(\omega) = \frac{\hat{s}(\omega)\hat{s}^*(\omega)}{2\pi} = \frac{|\hat{s}(\omega)|^2}{2\pi}$. There exists several efficient methods to compute it like the Fast Fourier Transform which requires the sampling frequency of the signal to be a power of 2 [46]. It is usually represented as a function of the frequency $f = \frac{\omega}{2\pi}$ and in decibels, $S_{dB}(f) = 10log_{10}S(f)$. The PSD is also more easy to interpret when it is smoothed by taking local averages over a short frequency band.

The two features which should be looked at with attention are the local peaks, indicating the oscillatory components of the trace coming from the input temporal structure or from internal properties of the cell, and the slope of the decay in log-representation. Many signals have a power spectrum behaving in a $\frac{1}{f^\alpha}$ fashion and α may give indications about the process underlying fluctuations of the signal. For a white noise, the spectrum is flat and $\alpha = 0$. For a Brown noise, as generated by a Wiener process, $\alpha = 2$ and fluctuations may be associated to a diffusive process. For pink noise, which can be generated by a shot noise process, $\alpha = 1$ and the origin of such fluctuations is still highly debated, a interesting hypothesis is that it could result from a self-organized critical process [47]. For more general Levy processes, α can take fractional value. It was shown in a recent study that different statistics of the visual input lead to different exponent in the scaling of the high frequencies power spectrum [48]. Anyway, these exponents reflecting power scale invariance should be considered with great care because their estimation is very sensitive on the frequency window considered. The PSD of cell X present a peak around 20 Hz and is otherwise nearly flat on the frequency window observed.

Wavelet analysis Fourier analysis describes in a compact manner the structure of temporal fluctuations in a signal but it would fail to detect transient oscillations, a solution can be to calculate the PSD over a time window for each point of time. Continuous wavelet analysis is another way to overcome this problem and to get a spectral representation of the signal at each time, a short introduction to this method is provided in [49] and advanced presentation can be found in [50]. It is gives spectral information at any point of time by convolving the signal with a family of wavelets of different temporal scales as shown on fig 1.3. The Morlet wavelets family, which will be used in the following, is generated by the mother wavelet

$$\Psi_{\omega_0}(t) = \pi^{-\frac{1}{4}} e^{-\frac{1}{2}t^2} e^{i\omega_0 t}$$

with

$$\Psi_{\sigma\omega_0}(t) = c_\sigma \pi^{-\frac{1}{4}} e^{-\frac{1}{2}t^2} (e^{i\sigma\omega_0 t} - \kappa_\sigma)$$

where $\kappa_\sigma = e^{-\frac{1}{2}\sigma^2}$ and $c_\sigma = \left(1 + e^{-\sigma^2} - 2e^{-\frac{3}{4}\sigma^2}\right)^{-\frac{1}{2}}$. There is a simple relation between wavelets and their mother, $\Psi_{\sigma\omega_0}(\frac{t}{\sigma}) = \sqrt{\frac{\delta t}{\sigma}}\Psi_{\omega_0}(\frac{t}{\sigma})$, with δt the time step of the signal. The wavelet transform is then $\tilde{s}_t(\omega) = \frac{1}{T}\int_T \Psi_\omega(t'-t)s(t')dt'$. It is actually simpler to use the Fourrier transform of this equation because the convolution becomes a simple multiplication. The Fourrier transform of the mother Morlet wavelet is $\hat{\Psi}_{\omega_0}(\omega) = \frac{1}{\pi^{\frac{1}{4}}} e^{-\frac{(\omega-\omega_0)^2}{2}}$ and the Fourrier transform for the rest of the family can be deduced by using the renormalization $\omega \leftarrow \omega' = \sigma\omega$ and $\hat{\Psi}(\omega') = \sqrt{\frac{2\pi\sigma}{\delta t}}\hat{\Psi}(\omega)$. The inverse FFT then gives the wavelets coefficients in an efficient manner. Transient oscillations appears as bump in the wavelets power spectra represented as a time frequency matrix, such a bump centered around 15Hz can be seen in fig 1.4 at 500ms, and those bumps could be detected automatically by using Gabor filters, see [51].

1.2.3 EEG

The electroencephalogram (EEG) is a very common signal in neuroscience, it can be recorded with an electrode at the surface of the scalp or with an intracranial electrode. As it is an analog signal, it can be processed with the same analysis as was presented for the membrane potential from which spikes have been removed. EEG signals are usually recorded on a longer period of time than the membrane potential with a sampling frequency around 1 kHz whereas the membrane potential is sampled at 10 kHz. Brain rhythms corresponding to different cognitive states can be tracked on this recording. Hans Berger recorded the first EEG signal on his son in 1929. He discovered the α-rhythm, an oscillation around 8 Hz in the occipital region of the brain associated to a rest state with closed eyes. It was further developed to study epilepsy and it is now widely used to measure the level of consciousness of patients or anesthesia depth with what is called the bispectral index. The functional role of these oscillations is

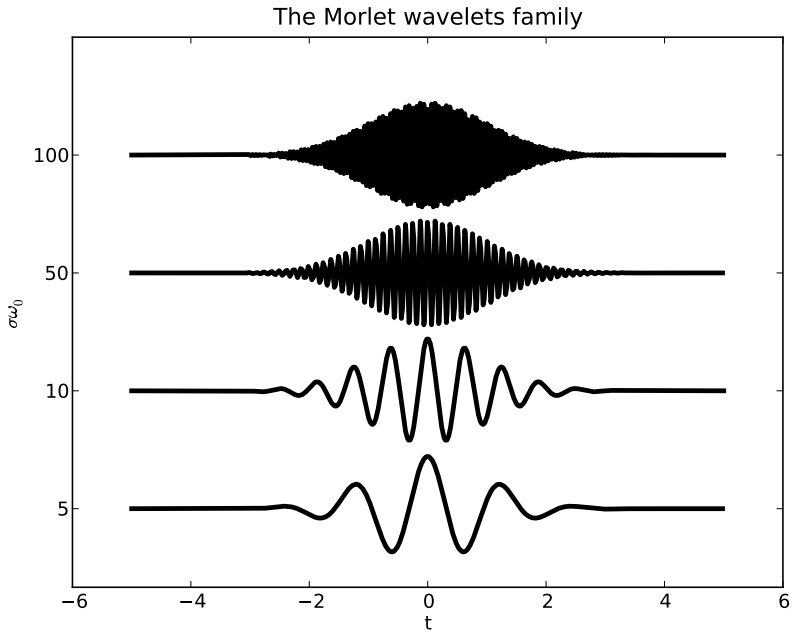


Figure 1.3: **The Morlet family** - Morlet wavelets at different scales.

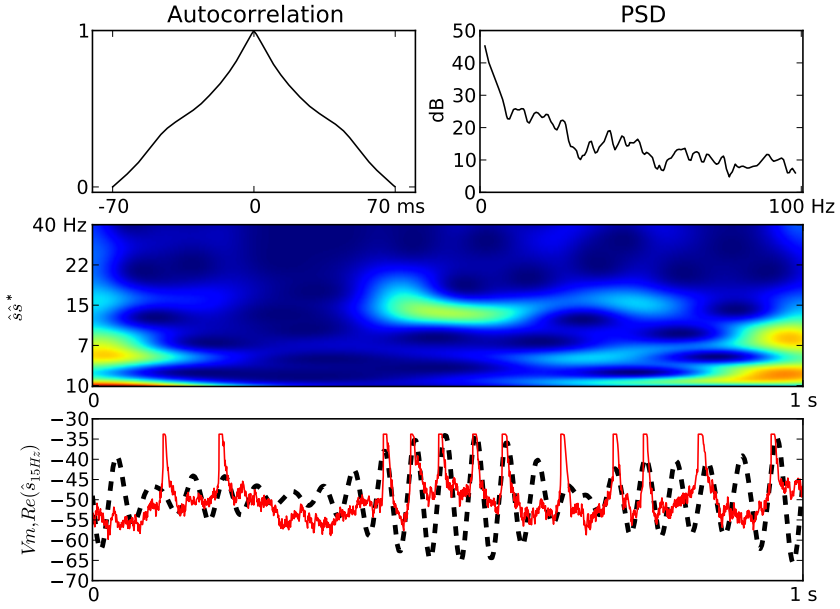


Figure 1.4: **Spectral properties** - (Top-left) Autocorrelation of the V_m trace. (Top-right) PSD of the V_m trace. (Middle) Time-frequency representation of the V_m signal. (Bottom) V_m trace (red) and 20 Hz component of the time-frequency representation (dashed).

still an active topic of research but the low frequency rhythms are usually associated to sleep or pathological states whereas cognitive processing is associated to higher frequency rhythms. The frequency bands can be summarized as:

Name	Frequency band	Functional role
δ	$1-3Hz$	Slow wave sleep
θ	$4-7Hz$	Memory retrieving
α	$8-11Hz$	Resting
β	$12-20Hz$	Attention
γ	$>20Hz$	Perceptual binding of a Gestalt

Recent research in cognitive neuroscience showed the importance of phase synchronisation between electrodes across brain areas [18]. The spatio-temporal structure of correlations between the 64 or 128 electrodes recorded makes it possible to discriminate between conscious and unconscious perception [52], it also reveals the attentional state of the subject [53].

The presence of brain rhythms makes the time frequency analysis particularly useful for EEG signals. For very long time series although, interesting events are difficult to capture and it is also difficult to infer temporal relationships between these rhythms. In the analysis described below, the signal is compressed and a graphical representation of the sequences describes the temporal organization of brain waves.

Example on an artificially generated signal. The artificial EEG Y, shown in fig 1.5, was generated by repeating 3 times the following sequence:

$$\delta \rightarrow \beta \rightarrow \beta+\gamma \rightarrow \theta \rightarrow \delta$$

with the γ oscillations are only active near the local maxima of the β oscillation. This sequence of transitions among rhythms and combinations of rhythms can be represented by a graph as shown on Fig??. The aim of the method proposed below is to extract the sequence of rhythms and combinations of rhythms activated and to build the graph corresponding to this sequence based on the time-frequency matrix.

Compression of the time-frequency matrix. The first step is to split the time-frequency matrix into blocks by choosing time and frequency intervals where the cutting are made. Regular sampling of the time at 1Hz enables a precise tracking of rhythms transitions and allows the detection of low frequency oscillations. For the frequency axis, the cutting can be based on the common frequency bands defined in the literature but it can also be adapted to the particular signal by taking frequencies of local minima of the spectrum as frontiers between the frequency bands. In the following, the frequencies are gathered in 4 bands ($b_1 = [1-8Hz]$:low frequency,$b_2 = [9-19Hz]$: middle frequency,$b_3 = [20-40Hz]$: high frequency and $b_4 = [41-100Hz]$: very high frequency). The locally integrated power spectral density with sampling window δt is obtained from the wavelet power density W by $L(t,f) = \frac{1}{\delta t}\int_t^{t+\delta t} W(t,f)dt$

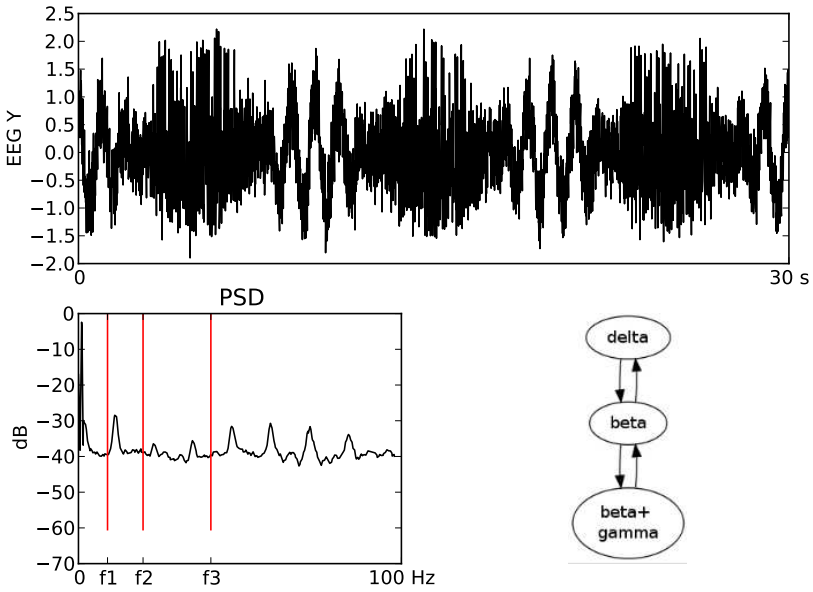


Figure 1.5: **Artificial EEG Y** - (Top) Artificial EEG Y. (Bottom-left) Power spectral density of the signal with limit frequencies of the 4 bands. (Bottom-right) State diagram representing the signal.

and the power density relative to the frequency band i is given by $B_i(t) = \frac{1}{f_{i+1}-f_i}\int_{f_i}^{f_{i+1}} W(t,f)df$. The compressed time-frequency matrix is then

$$\mathcal{C}_i(t) = \frac{1}{\delta t(f_{i+1}-f_i)}\int_t^{t+\delta t}\int_{f_i}^{f_{i+1}} W(t,f)df\,dt.$$

This compressed matrix will be used to detect transitions in the dynamics. It can also be used for an efficient online sonification of the signal where each frequency band code for a note with intensity given by the matrix values at each time. Transforming neuronal data into sound is useful because the human ear is very good at detecting temporal structure in audio signals.

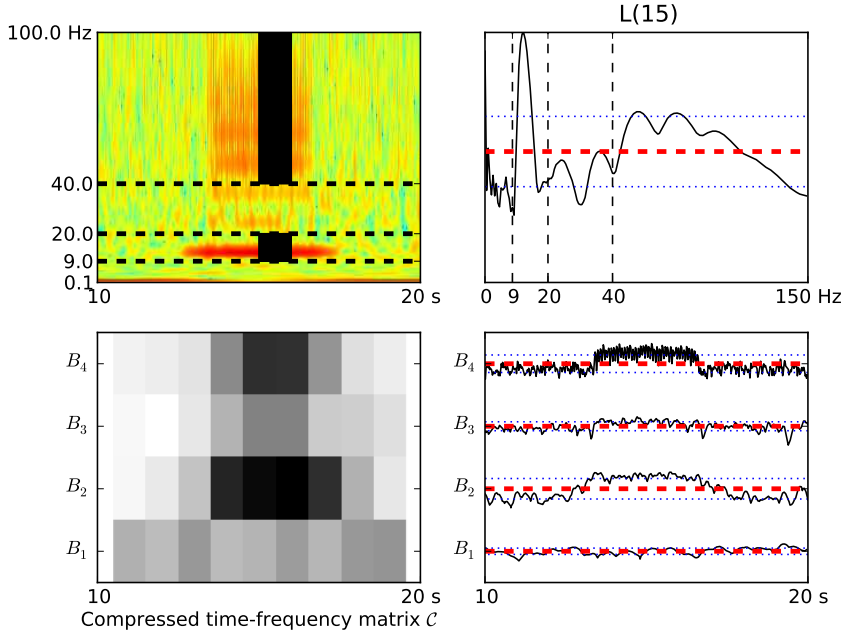


Figure 1.6: **Compression of the EEG** - (Top-left) Time-frequency representation of the signal, shaded areas represent activated bands and dashed lines represent frontiers of the frequency bands. (Top-right) Local power spectral density of the signal at t=15s. (Bottom-left) Compressed representation of the time frequency matrix. (Bottom-right) Dynamics of the integrated power in the four bands.

Definition of the symbols. Each column of the compressed matrix $\mathcal{C}$ provides a compact description of the frequency content of the signal at a time t. An empirical criterion $\theta_\epsilon(b_i,t) = (1-\epsilon)B_i(t) + \epsilon L(t)$ determines if a frequency

band b_i is activate at time t by

$$\text{if } \theta_\epsilon(b_i, t) > (1-\epsilon)E(B_i) + \epsilon E(L(t)),\ b_i \text{ is active}$$

. The band is temporally active at t when $\theta_0(t)$ is used as criterion, it has more power density than at other moments of time, and it is spectrally active if $\theta_1(t)$ is used, it has more power density than other frequency band. For intermediate values of ϵ, a frequency band is active depending on its power density relative both to other moments of time and frequency bands. In the EEG Y at $t = 15s$, considering the criterion θ_1, b_2 and b_4 are active. For each column of the compressed matrix, a 4-bits codeword $d_{b_4} d_{b_3} d_{b_2} d_{b_1}$ is formed based on the active bands of the signal. The digit b_i is equal to 1 if the frequency band b_i is active and 0 if it is inactive. The codeword for EEG Y at $t = 15s$ is 1010 and its decimal representation is 10. The same principle could be adapted to an arbitrary number of frequency bands and the codeword representation could be made more efficient by using Huffmann coding [7].

Building of the graph The signal can be represented as a string where each letter is the decimal translation of the codeword (between 0 and 15). The frequency of occurrences f of each letter and of each two letters word are then collected in a dictionary and a test is applied to each two letter word. If $f(ab) > f(a)f(b)$, the word ab is more frequent than it would be if a and b where appearing randomly in an independent way, the transition from a to b will then be reported on the graph. By this way, the graph of fig 1.5 for EEG Y is recovered. The result of this analysis for recorded EEG of 60 s duration is shown on fig 1.6. The detection of N-letters words can be made optimal by using Lempel-Ziv-Welch coding [8]. The graph of fig 1.7 is obtained from an EEG trace of 3 hours by drawing the strongest links. The graph can be used to build a statistical model like a markov chain giving the probability of occurrence of a state given the current. Transitions between brain states can also be represented as trajectories in a low dimensional phase space based on the spectral properties of the signal [54]. It would be interesting to check how these states relate to classes of neurodynamics at the single cell level.

1.3 Metrics and measures.

We consider a dataset $\mathbf{X} = (\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_n})$. Each data $\mathbf{x_i}$ is a p-dimensional vector representing a neuron recording. The neuron recording can be represented by its membrane potential trace, its spike train or p parameters extracted from those. We list below distances which can be used to evaluate the closeness of two data samples and measures representing the structure of the data set. We first investigate analog signals and then discuss the case of discrete data samples.

[7] Huffman coding is a way to perform loss-less compression of data by building a variable length code based on the probability of occurrences of the source symbols.

[8] LZW algorithm also performs loss-less compression. It is based on the encoding of substrings appearing in the data sequence to be compressed.

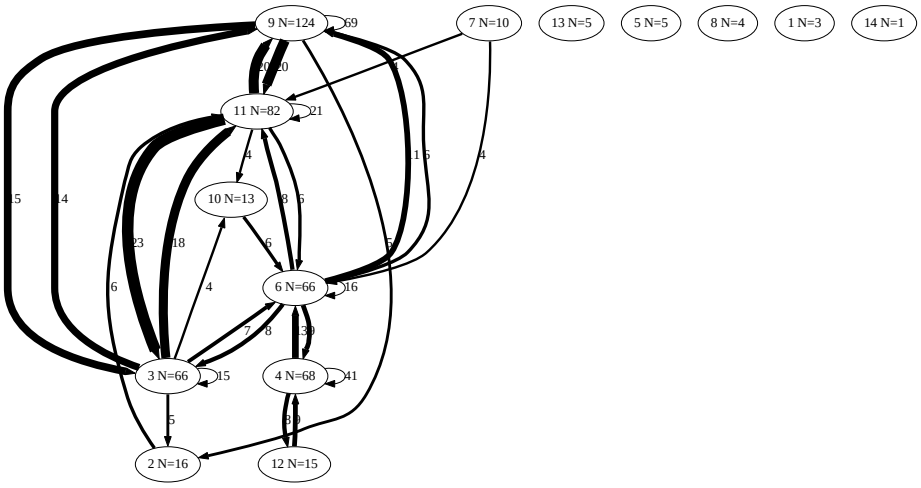


Figure 1.7: **State diagram of an EEG trace.** - Each node of the graph is a state with the number of occurrence N. The thickness of arrows represent the probability of transitions among those states.

1.3.1 Analog signals.

Classical distances. The Minkowski distance between two data samples depends on a parameter q, $d_q(\mathbf{x_i}, \mathbf{x_j}) = (\sum_{k=1}^{p} \|x_{ik} - x_{jk}\|^q)^{\frac{1}{q}}$. The Euclidian distance is the most natural metric to evaluate the similarity between 2 data samples. It is defined by $d_2(\mathbf{x_i}, \mathbf{x_j}) = \sqrt{\sum_{k=1}^{p} \|x_{ik} - x_{jk}\|^2}$. The city block distance is also used $d_1(\mathbf{x_i}, \mathbf{x_j}) = \sum_{k=1}^{p} \|x_{ik} - x_{jk}\|$. The distance matrix $D_{\mathbf{X}}$ of the dataset $\mathbf{X}$ is then obtained from the $d_{ij} = d_q(\mathbf{x_i}, \mathbf{x_j})$.

Correlation-based measures. The Pearson correlation coefficients are defined by $r_{ij} = \frac{1}{p} \sum_{k=1}^{p} \frac{(x_{ik} - \bar{x_i})}{\sigma_{x_i}} \frac{(x_{jk} - \bar{x_j})}{\sigma_{x_j}}$. It should not be confused with the covariance matrix, $Cov_{ij} = \frac{1}{p} \sum_{k=1}^{p} (x_{ik} - \bar{x_i})(x_{jk} - \bar{x_j})$. Other measures are defined in a similar way. The coherence of two signals is defined by considering the cross-correlation of the their power spectral density. The phase synchrony at specific frequency is obtained by cross-correlating the phase of these two signals at this band obtained from the time-frequency analysis.

1.3.2 Spike trains.

Pearson correlation. The simplest way to evaluate the similarity between two spike trains x_i and x_j is to consider their Pearson correlation coefficient defined similarly as that of a continuous signal. With such a measure, an exact synchrony of the two spike trains is necessary for being similar. For example, if B is just a copy of A with a shift δt greater than the time window used for the analysis, the correlation coefficient of A and B may be zero although the two

spike trains are very similar. Other metrics have been developed to avoid such pecularities.

Cost based method The Victor-Purpura distance [55] is based on the number of operations necessary for transforming x_i into x_j. The three basic operations considered are spike addition or deletion both having a cost of 1 and temporal displacement of δt having a cost of $\frac{\delta t}{\tau}$. The time constant τ is a free parameter of the defined distance.

Convolution based method As described above, a filtered version of the spike trains s_i and s_j are obtained by applying exponential or Gaussian kernels with width τ. A distance is then defined by [56]:

$$D^2(x_i, x_j) = \frac{1}{\tau} \int_0^T [s_i(t') - s_j(t')]^2 dt'.$$

For two spike trains differing only by the insertion or deletion of a spike, $D^2(x_i, x_j) = \frac{1}{2}$ and if the only difference is a shift δt of one spike, $D^2(x_i, x_j) = 1 - e^{-\frac{|\delta t|}{\tau}}$. Another similarity measure based on the filtered signals s_i, s_j is the following defined in [57]:

$$S(x_i, x_j) = \frac{\int_0^T s_i(t) s_j(t) dt}{\sqrt{\int_0^T s_i(t) dt} \sqrt{\int_0^T s_i(t) dt}}$$

. In both methods, a narrow width of the kernel makes the distance or similarity measure sensitive to spike jitter whereas with a broader width, the additional or missing spikes are detected.

Parameter free method Other methods for the estimation of (dis)similarity are described in [58]. The ISI-distance method has the advantage of being parameter free. The current interspikes interval is defined by $ISI_i(t) = min(t_{ik} | t_{ik} > t) - max(t_{ik} | t_{ik} < t)$ where t_{ik} is the k^{th} spike of the i^{th} neuron. The ISI-distance between x_i and x_j is then:

$$D(x_i, x_j) = \frac{1}{T} \int_0^T |I(t)| dt$$

with:

$$I(t) = \begin{cases} \frac{ISI_i(t)}{ISI_j(t)} - 1 & \text{if } ISI_i(t) \leq ISI_j(t) \\ -(\frac{ISI_j(t)}{ISI_i(t)} - 1) & \text{else} \end{cases}$$

1.3.3 Information theoretic measures.

Information theoretical measures, as an application field of probability theory, heavily relies on the estimation of the probability distribution of the data samples. As this estimation for finite size samples is often a difficult task, the following describes the concepts used in information theory with random variables and we provide simple application examples to illustrate it. For a more deep treatment of this subject, see [59], and for applications to spike train analysis see [40].

Shannon entropy

Definition and properties The Shannon entropy of a random variable X taking discrete values $\mathcal{X} = [x_0, ..., x_m]$, is $H(X) = \sum_{l=0}^{m} -P(X = x_l)log_2(P(X = x_l))$. H gives a measure of the uncertainty that is the number of yes/no questions it takes to guess the value of the random variable when following an optimal strategy based on the past occurrences of this variable. It is measured in bits and variables with maximal entropy for a given set $\mathcal{X}$ follows a uniform law. The Shannon entropy has the following properties:

- $H(X) > 0$
- $H(X, Y) = H(X|Y) + H(Y)$
- $H(X, Y) \leq H(X) + H(Y)$ with equality if and only if X and Y are independent.

H can be extended to continuous variables with the differential entropy, $h(p) = -\int_{-\infty}^{\infty} p(x)log(p(x))dx$ but classical properties of the entropy do not hold anymore. A more convenient way for the extension to continuous variables is to consider the relative entropy with a reference probability distribution q, also called the Kullback-Leibler divergence: $D_{KL}(p||q) = -\int_{-\infty}^{\infty} p(x)log(\frac{p(x)}{q(x)})dx$ where q is commonly taken as a Gaussian function. The differential entropy of a data sample of N points generated from a multivariate Gaussian law of average μ and covariance matrix $\mathbf{\Sigma}$ is $h(\mathcal{N}(\mu, \mathbf{\Sigma})) = \frac{1}{2}ln((2\pi e)^N|\mathbf{\Sigma}|)$ with $|\mathbf{\Sigma}|$ the determinant of the covariance matrix.

Estimation The estimation of differential entropy of a process is not an easy task because a precise estimation depends on the bin width used for estimating of the probability density. The entropy is thus bounded by $logN_{bin}$, the entropy of a random variable with uniform probability distribution having the same support. In fig 1.8, the entropy of a Gaussian signal at 10 kHz sampling frequency is estimated across time with the number of bins being 3 times the number of points in the signal used for estimation, the result is close to the theoretical value.

For cell X, the entropy of the membrane potential is compared with the entropy of a Gaussian variable with the same mean and variance in fig 1.8.

Before the first spike, the entropy increases linearly close to the behavior of a Gaussian random variable and it drops after the first spike. The entropy then grows at a much slower rate because there is a big part of the range (between -35 and -40 mV) which remains nearly unexplored.

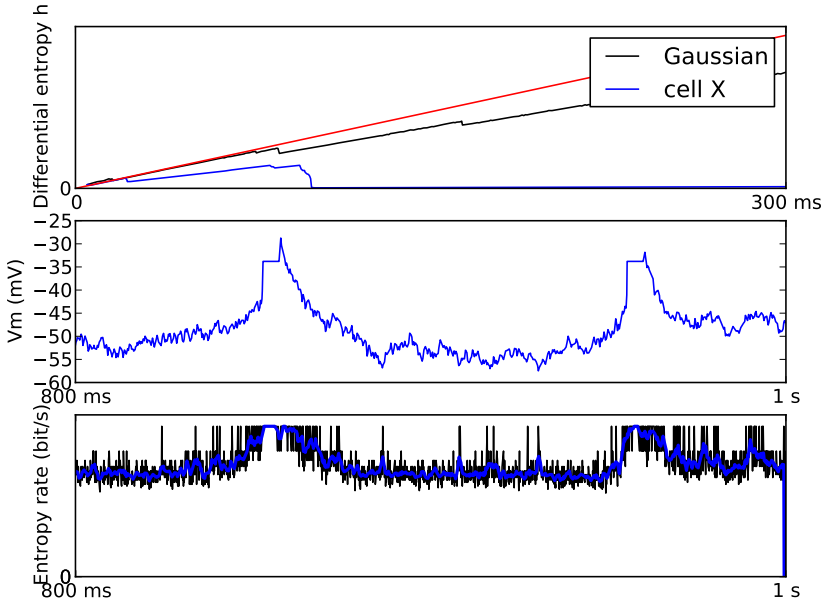


Figure 1.8: **Differential entropy and entropy rate** - (Top) Differential entropy estimated for a Gaussian process (black) and for the cell X (blue). The red line indicates the theoretical value for the Gaussian process. (Middle) 200ms of the V_m trace used for the estimation of the entropy rate. (Bottom) Entropy rate and its coarse grained version for 200 ms of the cell X.

Entropy rate The entropy estimate is difficult to interpret because it often far from its theoretical value and because its range drops drastically after a event like a spike occurrence. The entropy rate, $\frac{dh}{dt}$ is a better way to follows changes in the signal. As shown on fig 1.8, the rising part of a spike is associated with an entropy production and the falling part with entropy destruction.

Fisher information.

We suppose a parameter θ has to be estimated from observations of the random variable X. The likelihood function $f(X, \theta)$ gives the probability distribution

of X given θ. The Fisher information is then

$$\mathcal{I}(\theta) = \mathbf{E}[(\frac{\partial ln f(X,\theta)}{\partial \theta})^2 | \theta].$$

For sufficiently regular likelihood functions, it can also be written:

$$\mathcal{I}(\theta) = -\mathbf{E}[\frac{\partial^2}{\partial \theta} ln f(X,\theta) | \theta].$$

Applications. Based on this definition of information, the Cramer-Rao bound gives the limit of precision achieved by an unbiased estimator $\hat{\theta}$:

$$Var(\hat{\theta}) = \frac{1}{\mathcal{I}(\theta)}.$$

This theoretical bound can then be used for example to find the interval of confidence of the estimated frequency of an noisy oscillation. The Fisher information is also very important in probability theory because it is used to build a metric in spaces of probability distribution functions which is the starting point of information geometry [60].

Mutual information.

Definition and properties The mutual information between two discrete variables X and Y is defined from the entropy of the marginals and the joint probability distributions $I_2(X,Y) := H(X) + H(Y) - H(XY)$ or equivalently $I_2(X,Y) = H(X) - H(X|Y)$, it is symmetric $I_2(X,Y) = I(Y,X)$. There is no restrictions anymore to extend the definition to continuous variables with probability distributions p_X and p_Y and the integral version is $I_2(X,Y) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p_{XY} log(\frac{P_{XY}}{P_X P_Y})$, it is the Kullback-Leibler divergence between the joint law and the product of the marginal laws of X and Y. The mutual information measures the reduction of uncertainty in the estimation of X resulting from knowledge of Y. It is 0 for independent variables and it is $H(X)$ when Y is a copy of X.

Example on a multivariate Gaussian The 3 examples presented on fig 1.9 corresponds to sets of Gaussian variables (**X**,**Y**,**Z**) with the following covariance matrices:

$$A = \begin{pmatrix} .1 & .75 & .75 \\ .75 & .1 & .75 \\ .75 & .75 & .1 \end{pmatrix}$$

$$B = \begin{pmatrix} 1.22 & .7 & 0 \\ .7 & 1.22 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$C = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

In case A, where all variables are depending on each other, the mutual information is the same for any pair of variables. In case B, where only **X** and **Y** are correlated, the mutual information of (**X**,**Y**) is higher than for other pairs because observations on one of the variables reduces uncertainty about the other. When all variables are independent as in case C, the mutual information should be 0 for any pair but the finite size of the samples introduce a bias.

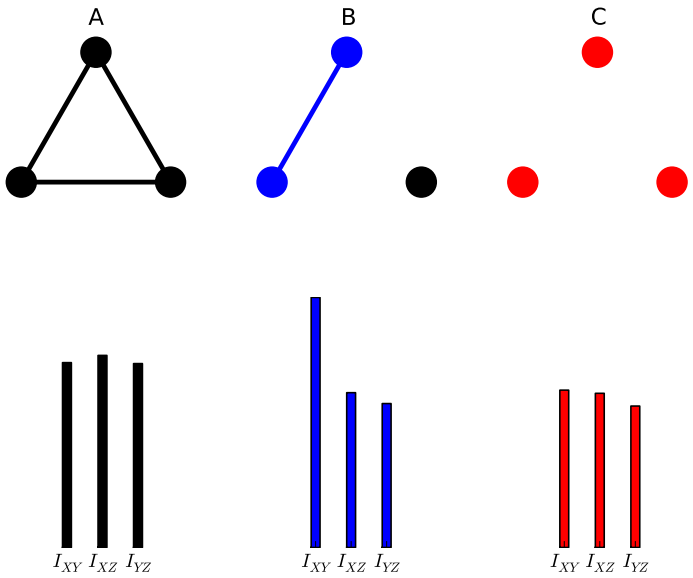


Figure 1.9: **Mutual information** - Mutual information for random processes generated by the multivariate Gaussian processes of covariance matrices A, B and C.

Neuronal complexity.

The mutual information can be generalized into the multi-information of any set of k random variables **X**: $I_k(\mathbf{X}) = \sum_{1<i<k} H(X_i)) - H(\mathbf{X})$, this quantity is also called the integration of the set and it is zero when all variables are independent. The neuronal complexity defined in [61] for a set of N variables

is:

$$\mathcal{C}_N = \frac{1}{N} \sum_{k=1,...n-1} (\frac{k}{n} I_N - < I_k >_k)$$

with $< ... >_k$ denoting an average over all the subsets of k elements. An approximation for weakly correlated variables is given in [62]:

$$\mathcal{C}_N = \frac{n+1}{24} (tr(R-I)^2 + tr(R-I)^3)$$

with R the correlation matrix [9] for off-diagonal elements. If the data $\mathbf{X}$ is generated by a coupled Ornstein-Uhlenbeck process [10]

$$d\mathbf{X_t} = \mathbf{X_t}(I - C) + \sigma d\mathbf{W_t}$$

, the complexity should be related to the coupling matrix. The previous approximation gives:

$$\mathcal{C}_N = \frac{n+1}{48} \sum_{i \neq j} (C_{ij}^2 + C_{ij}C_{ji}) + \frac{n+1}{96} \sum_{i \neq j \neq k} 3C_{ij}C_{jk}C_{ik} + \frac{n+1}{24} \sum_{i \neq j} C_i i (C_{ij}^2 + C_{ij}C_{ji}).$$

The neuronal complexity is thus related to the decomposition of the structure of the network in loops (first order term), 3-cycles (second order term),...The neuronal complexity thus quantifies how much a system is "more than the sum of its parts", a geometrical interpretation based on a comparison with families of exponential probability distributions can be found in [63]. The neuronal complexity is thus a promising measure for analyzing the huge amount of data arising from neuroscience experiments but it is still difficult to estimate it in an efficient fashion. There has been some recent progress for estimating the entropy of spike trains [64] and computational tools for this estimation are a growing field in neuroscience [65].

1.3.4 Attractor reconstruction

The signal recorded by intracellular electrodes or EEG devices is generated by non linear dynamical systems of high dimension but the effective dimension of the dynamics may be small due to the presence of rhythms. A theorem from Whitney and Takens further developed in [66] showed that for an attractor of effective dimension d, a delay-map in $\mathbf{R}^{2d+1}$ can be built which is qualitatively similar to the original attractor (that is there exists a diffeomorphism transforming one into the other). In this attractor reconstruction the delay and the dimension have to be chosen.

[9] The correlation matrix is composed of 1's on the diagonal and $R_{ij} = \frac{cov(X_i, X_j)}{\sqrt{var(X_i)}\sqrt{var(X_i)}}$

[10] Which can be considered as the linearization the stochastic Wilson-Cowan equations presented in Chapter 2

Optimal delay A rule of the thumb for choosing the delay of an oscillatory pattern is to take $\frac{1}{4}$ of the period. When there is no clear period in the signal, the optimal delay can be chosen as the minimum of the autocorrelation or the maximum of the mutual information $MI(\tau)$ between the signal and its time delayed version.

Correlation dimension A way to estimate the effective dimension of the attractor of a chaotic dynamical has been proposed in the 80's by Grassberger and Proccaccia in [67]. From the N points of the temporal signal $x_i = (y(i), y(i+\tau), y(i+2\tau), ..., y(i+k\tau))$ reconstructed from the original signal, the correlation sum is defined as:

$$C(r) = \frac{2}{N(N-1)} \sum_{i<j} \theta(r - |x_i - x_j|)$$

and the correlation dimension is $D = lim_{r\to 0} \frac{logC(r)}{logr}$ so that the correlation sum behaves as $C(r) \approx r^D$ for small r. This correlation dimension can be calculated for several values of the embedding dimension k and as k increases the optimal embedding dimension is obtained when D reaches a plateau. A public domain software called TISEAN [68] can be used for these calculations. For EEG signals, it has been suggested that the correlation dimension of EEG signals is reduced during sleep and pathological states like epilepsy [69].

1.4 Data classification

The previous sections showed that many parameters can be used to characterize signals corresponding to neuronal activity and that there are several ways to evaluate the similarity between two of those signals. In this section, we consider that some parameters have been extracted from the recordings and we wish to obtain a classification based on the comparison of these parameters. When many parameters are used, it is difficult to perform a efficient classification. This "curse of the dimension" can be attenuated by reducing the dimension of the parameter space.

1.4.1 Preprocessing of the data set

Normalization of the feature space The dataset $\mathbf{X}$ is first normalized by $\tilde{\mathbf{x_i}} = (\frac{\mathbf{x_{ik}} - \overline{\mathbf{x_k}}}{\sigma_{x_k}})_{1<k<p}$. After this operation, all parameters have the same variance 1 and the classification based on this normalized dataset is not affected by the range over which the parameter take values.

Orthogonalization of the feature space (PCA) Principal components of the dataset are extracted using the covariance matrix $C = \tilde{X}^t\tilde{X}$. C is symmetric so it can be diagonalized $C = {}^tP\Lambda P$. Λ is a diagonal matrix where each diagonal term represent the contribution of the corresponding eigenvector to the total

variance. Vectors are then reordered from the one with the biggest eigenvalue to the one with the smallest eigenvalue. For the classification of the recordings, the dimension of the parameter space can be reduced by selecting only the M first vectors explaining 90% of the variance.

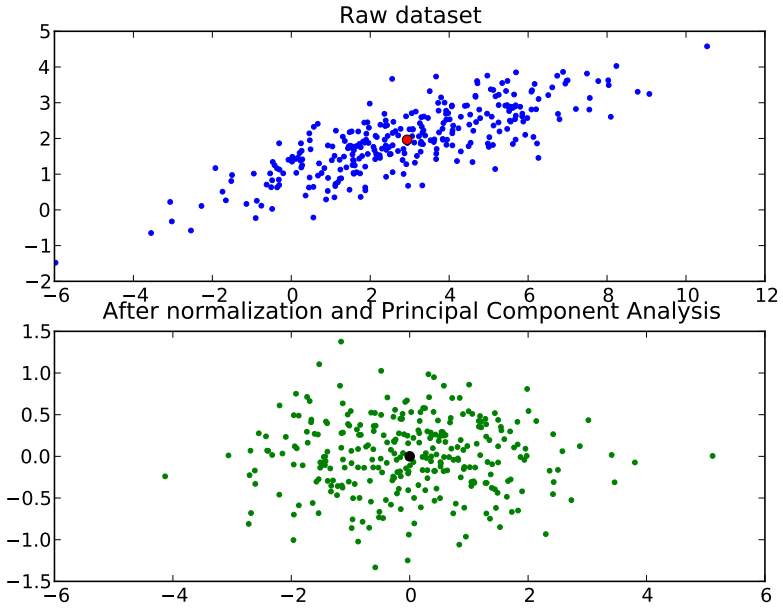


Figure 1.10: **Preprocessing of a 2D-Gaussian dataset (300 points)** - (Top) Raw data set. (Bottom) After PCA, the principal axis of the Gaussian becomes aligned with the horizontal axis.

1.4.2 K-means clustering

Description of the algorithm.

The K-means method is a way to clusterize cells by making an a priori assumption on the number of clusters K [70]. We will discuss possible ways to select seeds and generate partitions of the parameter space. This method is simple and efficient, it is widely used in the scientific community but it also have pitfalls of all unsupervised learning method. A common example of application where it gives a poor result is the Fisher iris data base. We consider $\mathbf{X}$, a set of n data points $(x_i)_{1\leq i\leq n}$ in $\mathbb{R}^p$. The algorithm will partition the points around K centers $(C_k)_{1\leq k\leq K}$ minimizing a potential function $\phi = \sum_{1\leq k\leq K} min_{x_i\in C_k}\|x_i - c_k\|^2$. This potential function is monotonically decreasing during the K-means algorithm and it will always terminate because the number of possible partitions is

bounded by K^n. Although the clustering procedure will always terminate, finding the globally optimal partition is a NP hard problem [11] and we will discuss possible solutions to approach this global optimum.

Seed selection. Seeds are the initial cluster centroids $(c_j^0)_{1<j<k}$ and the simplest way to select it is to choose randomly K points as seeds with uniform probability law from the data set. This method is standard but better results can be obtained with a careful seeding as shown [71]. The selection of the seeds is the following where $D(\mathbf{x})$ is the distance of the point $\mathbf{x}$ to the closest centroid already chosen:

a Take the first centroid c_1 randomly with uniform probability law.

b Take next centroid among $\mathbf{X}$ with probability $\frac{D(\mathbf{x})}{\sum_{\mathbf{x}\in\mathbf{X}} D(\mathbf{x})^2}$.

c Repeat a and b until K centroids are selected.

For this Kmeans++ algorithm, the potential function is shown to check $\mathbf{E}[\phi] < 8(lnK + 2)\phi_{OPT}$ where ϕ_{OPT} is the optimal partition of the data set.

Iterative procedure. The process unfolds in two steps repeated until convergence is obtained:

1 Attribute each data point to its closest centroid by computing $j_c(\mathbf{x_i}) = min_j d(\mathbf{x_i}, \mathbf{c_j^t})$. The j corresponding to the minimum distance is the cluster id which will be attributed to the cell.

2 Compute the new centroid position $\mathbf{c_j^{t+1}} = (\sum_{x_i k \in C_j^{t+1}} x_{ik})_{1\leq i\leq p}$.

3 Steps 1 and 2 are repeated until successive centroids stay close to each other, $d(\mathbf{c_j^t}, \mathbf{c_j^{t+1}}) < \epsilon$.

where t indicates the iteration of the process.

Selection of a "correct" partition.

The procedure described above is repeated many times and a criterion has to be defined so that the partition is considered as robust or not too "bad". As we already saw the result of the K means depends on initial conditions, so those are changed at each iteration. The distance to used also affect the resulting partition, the Euclidian distance is commonly used for K-means procedure and the city-block distance is used when medians are taken as centroids.

[11] NP hard problems take a very long time to solve when the size of the system grows.

Dunn Index We can consider a partition to be "correct" if its clusters are sufficiently compact and well separated. The relevant measures for this are the radius of the largest cluster, $R_{max} = max_{C_k} \frac{\sum_{x_i \in C_k} \|x_i - c_k\|}{|C_k|}$ and the minimum distance between clusters, $L_{min} = min_{C_i, C_j} \|c_i - c_j\|$. These two quantities are combined in the Dunn index $DI = \frac{L_{min}}{R_{max}}$. Among the partitions generated by the K-means algorithm, we select the one with the highest DI and we stop the selection when there is no improvement in the DI. We also use the DI to detect optimal K for the partition.

Frequency of occurrence of the partition After many runs of the K-means algorithm, the final partition will often be the same. A way to quantify this is to run the algorithm many times and to take the partition which is most often encountered in this process.

Example on a mixture of Gaussian distributions To illustrate the K-means algorithm, we generated artificially two data clouds **A** and **B**. In the data set **A**, 180 points are randomly chosen following a 2D-Gaussian probability distribution function (pdf). The data set **B** is prepared with a mixture of Gaussian probability distribution functions where for each Gaussian, having different means and covariance matrices, 30 points are chosen randomly. The data set **A** lack of any internal structure and the frequency of the most often encountered partition in 10000 repetitions of the K means algorithm as a function of K, the number of centroids considered, decreases montonically in an exponential fashion. In the data set **B**, deviations from this monotonical decrease shows the non-homogeneity of the data set and the drastic drop when K goes from 6 to 7 suggest that the data set can be well represented as a collection of 6 clusters.

1.4.3 Tree Building.

The K-means method to find clusters in a data set is stochastic, because the final partition depends on the initial conditions. All the more, it is a "flat" method because the obtained clusters are disjoint. A clustering procedure is hierarchical if in the resulting partition, each cluster is formed with subclusters, themselves containing subclusters,...If two points are grouped together at a given level, they will stay grouped at higher levels. The natural representation for a data set on which hierarchical clustering has been applied is a tree, also called a dendrogram. There are two possible ways to perform hierarchical clustering, it can be started with every data sample in a singleton cluster and this bottom-up process is an agglomerative tree building, or it can be started with a giant cluster containing all the data samples splitted successively until each cluster contains only one data sample and this top-down process is a divisive tree building, both methods are described in [70] and we here focus on the agglomerative method.

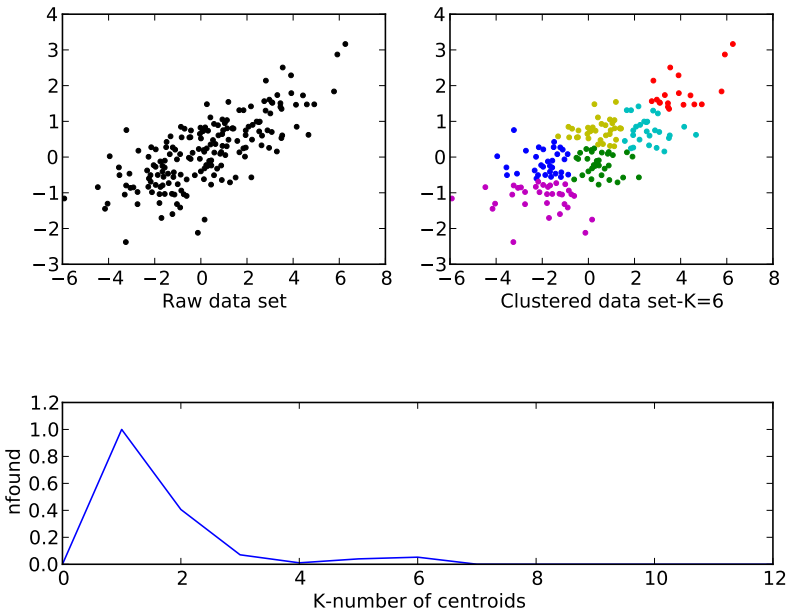


Figure 1.11: **Data set A** - 180 points chosen randomly with a Gaussian pdf.

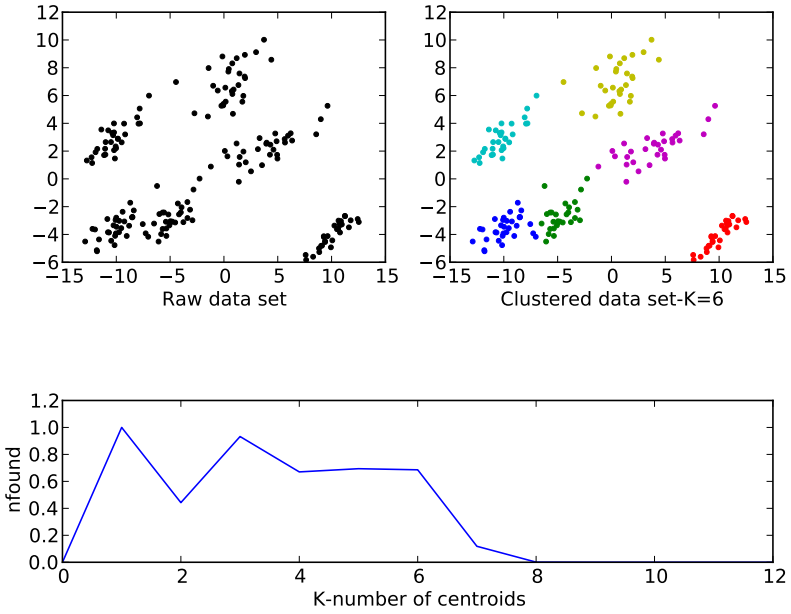


Figure 1.12: **Data set B** - 180 points chosen randomly with a mixture of 6 Gaussian pdfs.

Agglomerative tree building.

Algorithm. In the agglomerative tree building method, every data point is considered as a cluster or node. The process then merges the two closest clusters and iterates until only one cluster remains. A pairwise distance has to be chosen so that we can evaluate the distance between two cluster. The pairwise average-linkage (pal) distance considers the distance between two clusters as the average over all the pairwise distances between elements of the two clusters. The process consists of the following steps:

1 Merge the closest clusters $C_i, C_j = argmin_{i,j} d_{pal}(C_i, C_j)$ into the cluster C_l.

2 Repeat 1 until C_l contains all elements of the data set.

Application to Gaussian mixture. As for the K means clustering, the Tree Building algorithm is applied to the data set A, 180 points randomly chosen from a mixture of 6 Gaussian distributions. By cutting the tree at an appropriate depth, 6 clusters are obtained corresponding to the 6 Gaussian distributions from which the data samples are generated.

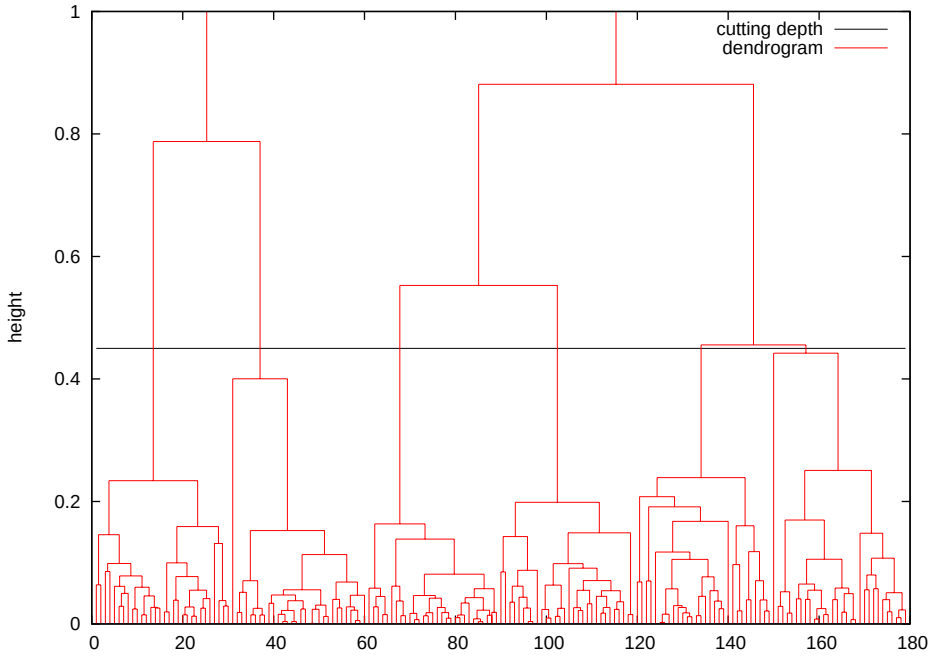


Figure 1.13: **Data set B** - 180 points chosen randomly with a mixture of 6 Gaussian pdfs.

Distance matrix and cutting of the tree

Base of tree At the base of the tree, cells are ordered in a way related to the hierarchical structure of the dataset. This order can be considered as an interesting way to enumerate the data samples. In the distance matrix of fig 1.14, the lower triangle stands for the data samples in their initial order and the upper triangle for the order resulting from the tree building procedure. The 6 clusters are clearly detectable in the upper triangle.

Cutting the tree An arbitrary number of clusters (bounded by the number of cells) can be obtained by cutting the tree at an appropriate depth or more elaborate cutting methods [72]. By splitting the tree into the same number of clusters as resulting from the K means procedure, the two partitions can be compared. A possible way to compare the partitions obtained from two different methods or in two different experimental conditions is to build a matrix H of dimensions (K_1, K_2) where K_1 is the number of clusters resulting from the first method and K_2 for the second. Elements of the matrix are filled as follows:

$H_{i,j} = |$data samples belonging to cluster i by method 1 and j by method 2$|$.

1.4.4 Kohonen network.

The self-organizing map algorithm is a biologically inspired model used to map data samples $(\mathbf{x_i})_{1<i<n}$ from the input space to nodes $(y_j)_{1<j<k}$ of the feature space. A weight vector $\mathbf{w_j}$ is associated to each node and a dynamic evolution of these weights representing learning, leads to a low dimensional representation of the data samples. The simplified version presented below is often referred to as a Kohonen network [73] and more sophisticated models of self-organizing maps will be described in the Chapter 3 dealing with models of V1 formation. The relaxation time of this dynamic evolution is a free parameter of the algorithm. The feature space is often taken on a 2D regular grid of dimensions (N_x, N_y). In this algorithm, the final result is dependent on the order of presentation of data samples.

SOM algorithm For each data sample presentation, a competition is taking place and the winner dictates the weights evolution dynamics in its neighborhood. The process is as follows after random initialization of the weight vectors:

1 Compute the activation for each node in the feature space $y_k = \sum_{(1<j<p)} |w_{kj} - x_j|$ and select the closest one y^* from the data sample x.

2 Update weights according to the following learning rule:

$$\mathbf{w_k}(i+1) = \mathbf{w_k}(i) + \alpha(i) h_{*k}(i)(\mathbf{x} - \mathbf{w_k}(i))$$

.

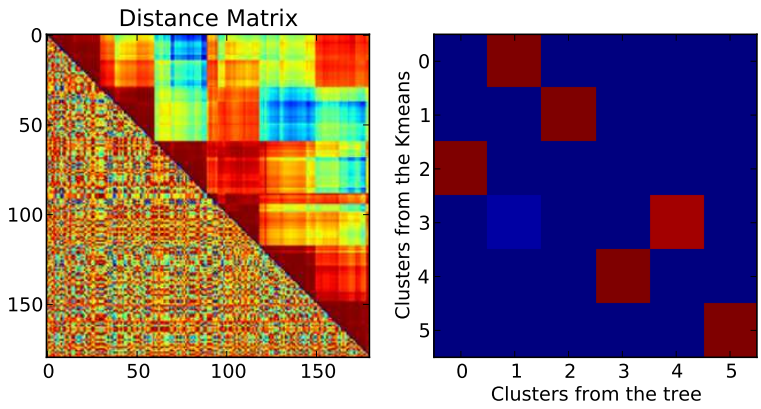


Figure 1.14: **Similarity and comparison of the obtained partitions** - (Left) The lower triangle shows the similarity in a random order and the upper triangle shows the similarity between data samples ordered according to the tree. (Right) The quantity in (i,j) indicates how many cells fall into cluster i with the tree partition and into cluster j with the K means partition.

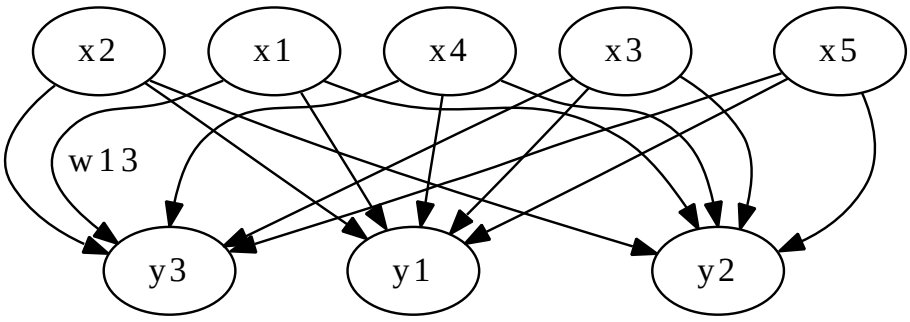


Figure 1.15: **SOM network.** - The network is composed of input nodes x and map nodes y.

3 Renormalize the modified weights so that $\|\mathbf{w}\| = 1$.

The learning rate $\alpha(i)$ is a decreasing function of time depending on the relaxation time τ, a common choice is $\alpha(i) = \tau(1 - \frac{i}{n})$. The neighborhood function $h_{*k}(i)$ is 1 if the node k is closer than the influence radius $R(t)$ from the winner and 0 if it is farther. The influence radius decreases as $R(i) = R_{max}(1 - \frac{i}{n})$ with $R_{max} = \sqrt{N_x^2 + N_y^2}$.

Biological interpretation. The step 1 can be seen as an implementation of a neural field and the step 2 is an example of implementation of a plasticity rule. The algorithm is thus inspired by cognitive theories about assosiative memory and adaptive learning. A more detailed model inspired from the SOM but including recurrent connections will be studied in Chapter 3. The relaxation time τ is a free parameter of the model and it should be adapted to the studied dataset.

Application to the data set B The SOM algorithm is applied to the data set B with $\tau = 0.02$ and 200 iterations on 5x5 grid. As shown on Fig 1.16, the data samples are composed of 6 main clusters with more than 15 data samples and a collection of smaller clusters. An important property of the SOM algorithm is that close points in the data cloud will fall onto close points of the map.

1.4.5 Misclassification and metasimilarity

Each method employed to partition a data set should be related to the question asked by the analyst and its advantages and pecularities should be well understood. On the one side, the K-means method is an easy way to determine the number of clusters into which the data set will be split. A good K can be deduced from the evolution of the Dunn index or the frequency of occurrence of the partition most often encountered. On the other side, it offers no indication about the relation between clusters. The tree building is very nice to visualize the fine hierarchical structure of the data set and to provide an order in accordance to this hierarchical structure but, as a deterministic method, its result is highly affected by outliers. All the more, it is often difficult to know at which depth the tree should be cut to give a good partition. To have an idea of topological relations between clusters, the SOM algorithm makes a very good job but it necessitates a tuning on several parameters (relaxation time, number of iterations, size of the grid).

By employing multiple techniques, the resulting partitions can be compared. In fig 1.14, the cluster C_5 of the partition P_{KM} (obtained from K-means) is the same as the one from the partition P_{TB} (obtained from tree building). The cluster C_0 in P_{KM} becomes the cluster C_2 in P_{TB}. A data sample which is in cluster C_3 in P_{KM} lands at C_1 in P_{TB} whereas its expected destination is C_4.

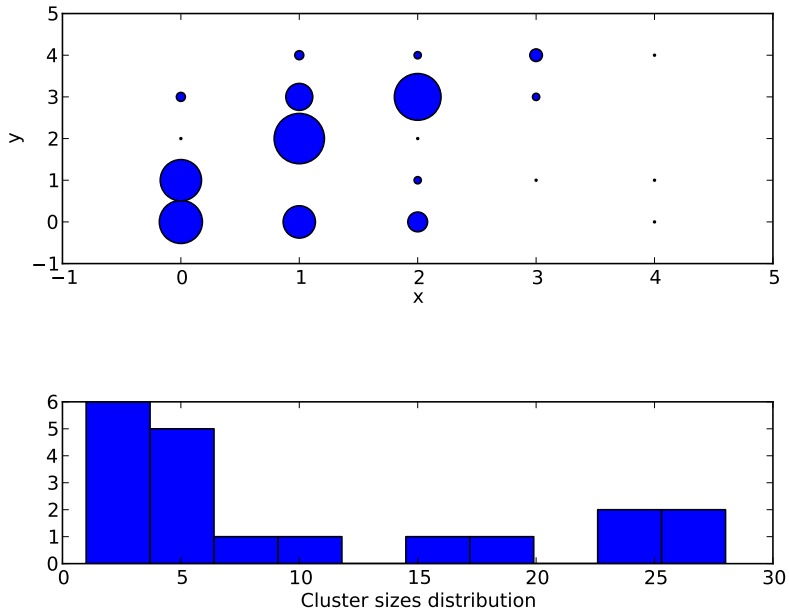


Figure 1.16: **SOM classification of the data set B** - (Top) Resulting 5x5 SOM. (Bottom) Cluster size distribution.

The matrix H thus provides an easy way to detect inconsistent classification. These misclassified points are near the frontiers between different clusters.

The 3 classification presented can be used to formalize a new notion of similarity. The metasimilarity between the data samples i and j is defined by

$$\mu_{ij} = 1 + \delta_{ij}^{KM} + \delta_{ij}^{TB} - d_{ij}^{SOM}$$

with $\delta_{ij}^{KM} = 1$ if the data samples i and j are in the same cluster of P^{KM}, $\delta_{ij}^{KM} = 1$ if the data samples i and j are in the same cluster of P^{TB} and d_{ij}^{SOM} is the Euclidian distance between clusters of the data samples i and j in the SOM normalized between 0 and 1.

If $\mu_{ij} = 3$, the data samples i and j lands in the same cluster whatever the clustering method and in that sense data samples i and j are metasimilar. If $2 \leq \mu_{ij} < 3$, the two data samples are simililar but the cluster to which they belong could be split into subclusters given by the SOM to highlight the difference. If $1 \leq \mu_{ij} < 2$, i and j belongs to different clusters or there may be a misclassification problem for one of the two samples. Finally, if $\mu_{ij} < 1$, the two samples are clearly in different clusters. The metasimilarity thus formalize in a comprehensive way the results of different data classification methods.

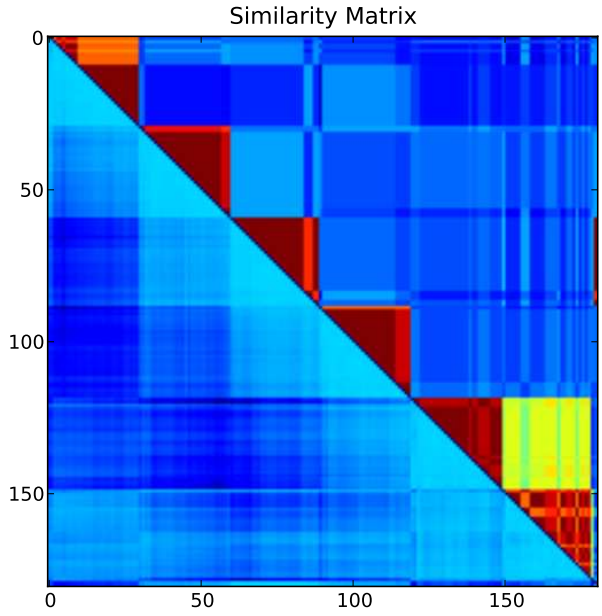


Figure 1.17: **Similarity and metasimilarity** - The lower triangle shows the similarity and the upper triangle shows the metasimilarity between data samples.

1.5 Application to electrophysiological recordings.

We consider the membrane potential recorded with an intracellular electrode in area 17/18 of the anesthetized cat (alfatesin). Some parameters are extracted from electrophysiological recordings to build the data sets. A first data set, Spt150, is composed of 150 data samples from cells recorded during ongoing activity. Another data set, Vis143, is composed of 143 data samples from cells recorded during the presentation of a visual stimulus. The obtained classification is compared to the same pool of cells during ongoing activity, Spt143.

1.5.1 Parameters extracted from electrophysiological recordings.

We show in fig 1.18 the list of 25 parameters with their average and standard deviation over the population.

There are 3 groups of parameters: parameters related to the distribution of the membrane potential, parameters related to the spectral properties of the membrane potential and parameters related to spikes of the cell. In parameters related to the distribution, we find moments of the distributions and coefficients reflecting the asymmetry or the deviation from unimodality. Most of spectral parameters are fractions of the power spectrum integrated over a frequency band. The instantaneous firing is the reciprocal of the interspikes interval whereas the average firing rate is the spike count divided by the recording duration.

1.5.2 The on-going activity data set (150 cells).

Optimal partition. A home made K-means method is applied to Spt150 with values of K from 2 to 9 and with the Dunn index as an optimization criterion. Only the 9 first components of the PCA are necessaty to explain 90% of the variance. The light blue curve of fig 1.19 corresponding to K=6 shows rapid convergence to its optimal Dunn index. Moreover, the optimal value for K=6 (black curve) is higher than the optimal value for K=5 (red curve). This suggest that the partition with K=6 gives a better description than with other values of K. The optimal partition for K=6 is given on fig 1.20 and the successive splittings of the data cloud can be visualized on fig 1.21. Note that the clustering obtained in fig 1.20 is different from that of fig 1.21 for K=6 because k-means++ method was applied to choose initial conditions in the first case whereas it is taken randomly with uniform probability in the second case.

Description of the partition. By a multifactorial decomposition analysis in Matlab, the main relevant parameters for the description of the partition are those related to the distribution of V_m and the frequency content in high frequency. The largest cluster (red) is composed of cells having a Gaussian distribution. The smallest cluster (pink) is composed of cells having a symmetric

Parameter	**Average**	**Standard deviation**
Mean of V_m [mV]	-66.94	3.91
Std of V_m [mV]	4.54	1.60
Skewness of V_m	0.61	0.53
S1	1.38	1.34
S2	0.40	0.40
kurtosis of Vm	0.34	1.42
Dip Test	4.21	12.66
Separability	22.32	15.38
Discretness	82.69	9.81
Regression coefficient		
from a fit of the distribution with a Gaussian	0.93	0.084
Regression of 2 Gaussians - Regression 1 Gaussian	0.054	0.074
Regression of 3 Gaussians-Regression of 2 Gaussians	0.0029	0.010
Slope coefficient of the Vm PSD	-2.74	0.43
Power ratio	0.93	0.44
Power in the delta band of the V_m PSD [dB]	0.41	15.86
Power in the theta band of the V_m PSD [dB]	30.83	7.79
Power in the alpha band of the V_m PSD [dB]	16.32	5.07
Power in the beta band of the V_m PSD [dB]	12.22	5.10
Power in the gamma band of the V_m PSD [dB]	14.40	11.09
Ratio between maximal autocorrelation		
and mean autocorrelation of V_m	3.87	2.71
Relaxation time of the autocorrelation [ms]	35.92	24.88
Average firing rate [Hz]	5.78	5.84
Instantaneous firing rate [Hz]	34.40	30.36
Coefficient of variation of interspikes intervals	1.28	0.40
Slope coefficient of the ISI distribution	-0.13	0.50

Figure 1.18: Parameters used for the classification of the cells.

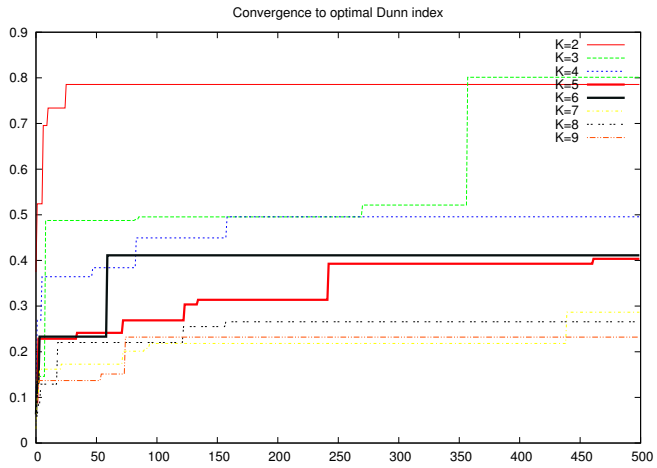


Figure 1.19: **Dunn index for Spt150** - The evolution of the Dunn index over iterations is plotted for several values of K. The black curve corresponding to K=6 saturates faster and at a higher value than the red curve corresponding to K=5.

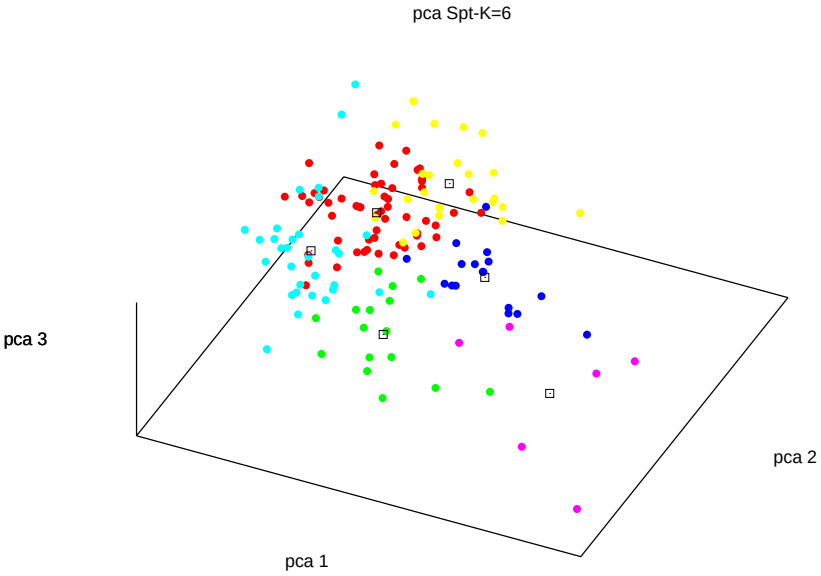


Figure 1.20: **Optimal partition of Spt for K=6**- Each cluster is represented by a different color and centroids are represented by square boxes. Axes are the 3 first principal components.

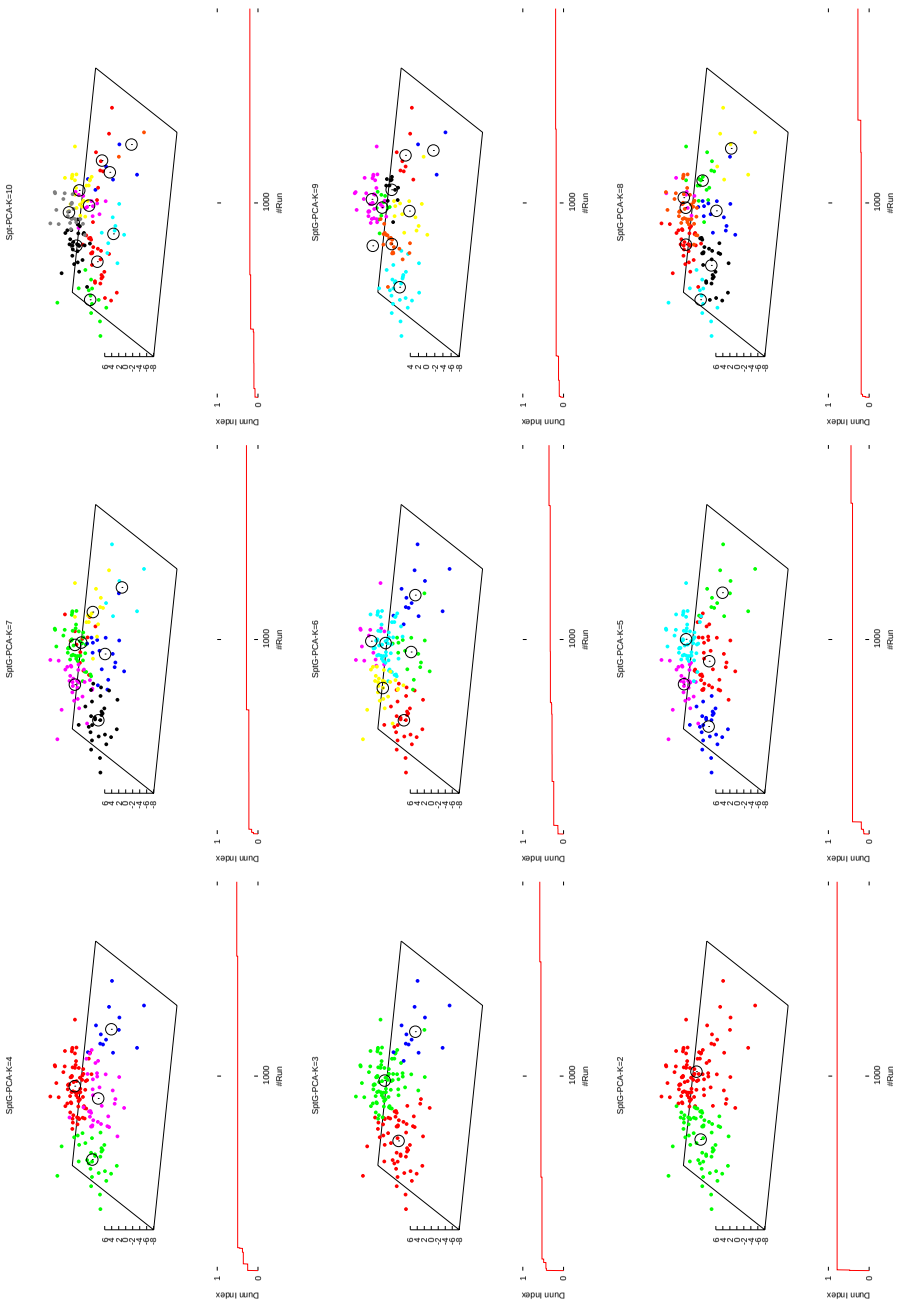


Figure 1.21: **Clustering of Spt150 for different values of K** - For each K, the clusters are represented by different colors in the PCA space. Red curves show the Dunn index evolution over 2000 iterations.

bimodal distribution and up and down transitions are correlated with the EEG fluctuations. The yellow cluster corresponds to very sharp distribution of V_m, that is cells with slow variations of V_m have correlations with the EEG. The deep blue and green clusters correspond to cells having asymmetric bimodal distributions dominated by the up and the down state respectively. The light blue cluster correspond to cells having a broad V_m distribution but only one peak. The data samples which are the closest from the centroids give a good summary of this partition and wo examples of membrane potential for each cluster with the associated EEG signal are shown on fig 1.22.

1.5.3 The visually evoked data set (143 cells).

For the visually evoked data set of 143 cells, Vis143, 11 components of the PCA are necessary to explain 90% of the variance. As can be seen on fig 1.23, the tree obtained for this data set is not well equilibrated because some few cells behaves very differently from the major part of the cells. This can be checked on the H matrix of fig 1.24, with all cells gathering in the fourth cluster of the tree based partition. The metasimilarity matrix describes the partition at a finer scale.

Clusterization of Spt143. The same algorithms were applied to the reduced data set of 143 cells of on-going for which the visually evoked activity is available. The K-means with the frequency of occurrence of the partition as an optimization criterion gives 4 clusters as optimal partition. Similarly to the clusterization of Vis143 cutting the tree in 4 clusters gives a poor result because it gathers most of the cells in a giant cluster.

Visually evoked activity compared to the spontaneous activity In the first three components of the PCA, the standard deviation is 1.25 for the spontaneous activity whereas it is 1.11 for the visually evoked activity. The visual stimulation thus pushes the activity toward the central red cluster of fig 1.27 corresponding to a Gaussian distribution of the membrane potential. There is no creation of a new domain for the dynamics as shown on fig 1.27. The comparison matrix for the clustering of the spontaneous activity data set and the visually evoked data set shows that there is a correspondance between the cluster 3 of the spontaneous activity data set and the cluster 1 of the visually evoked data set and another between cluster 2 of Spt and cluster 3 of Vis. Thus for clusters 2 and 3 of the spontaneous activity, cells don't jump to other clusters but stay close together when a visual stimulus is presented.

1.6 Conclusion

We presented a method to characterize and classify neuronal dynamics. Each classification has its own pitfalls and it is thus necessary to rely on a multi-algorithm approach to obtain a robust classification. We now summarize the classes obtained in the K means partition. The red cluster in fig 1.20 is the

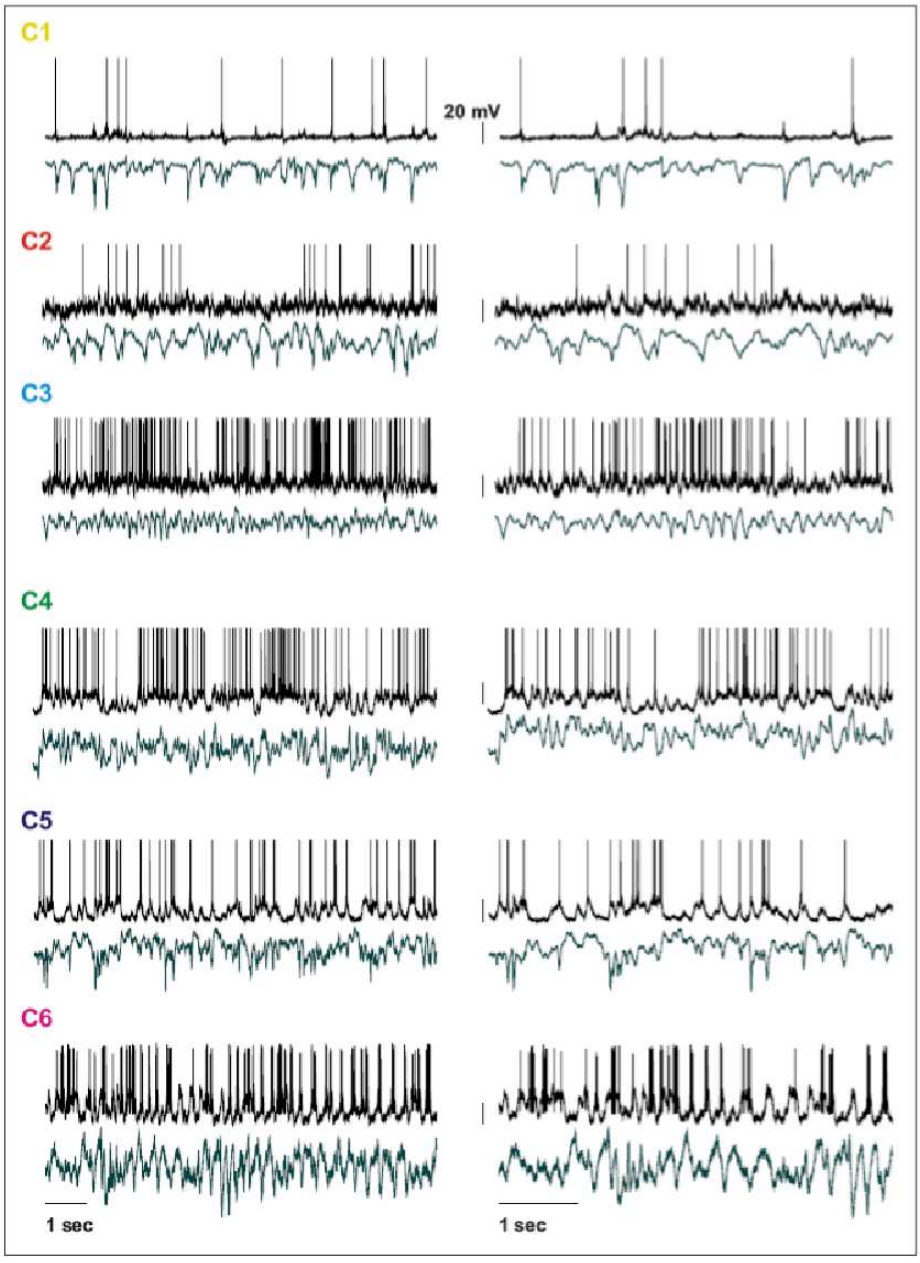


Figure 1.22: **Example of dynamics from the 6 clusters.** - For each cluster, the upper traces are membrane potentials and below is the EEG.

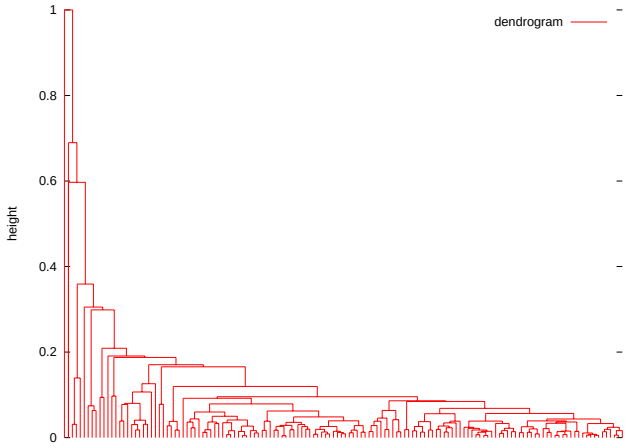


Figure 1.23: **Tree obtained for the 143 cells of the visually evoked activity dataset.**

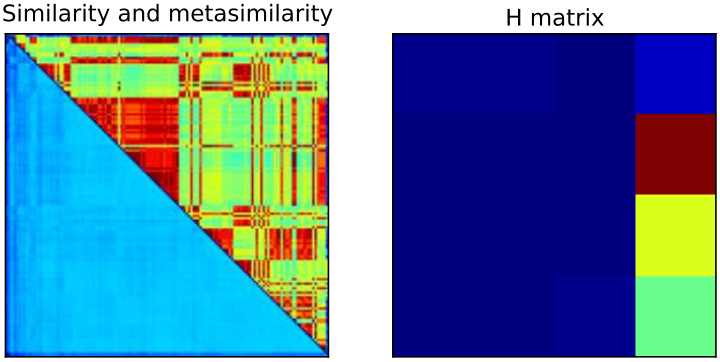


Figure 1.24: **Summary of the clustering algorithms for the 143 cells of the visually evoked activity data set.** - (Left) The lower triangle is the similarity matrix and the upper triangle is the metasimilarity matrix. (Right) H matrix obtained from the K-means partition and the cutting of the tree into four components.

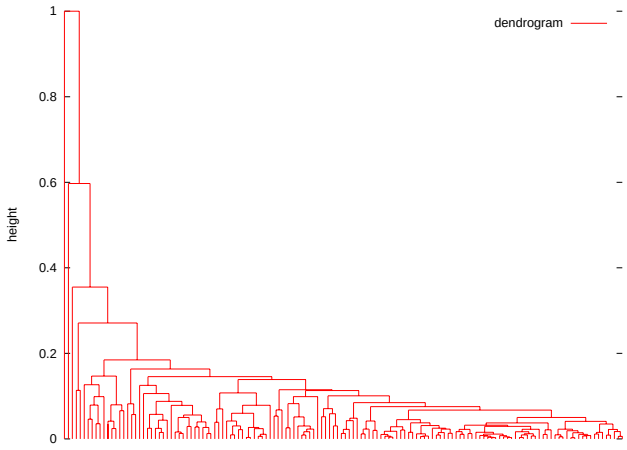


Figure 1.25: **Tree obtained for the 143 cells of the spontaneous activity dataset.**

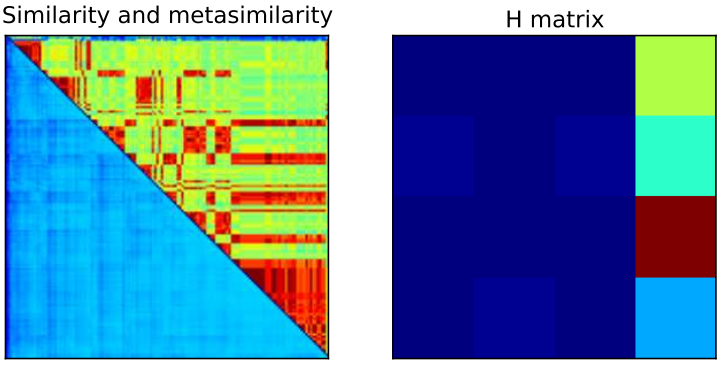


Figure 1.26: **Summary of the clustering algorithms for the 143 cells of the spontaneous activity data set.** - (Left) The lower triangle is the similarity matrix and the upper triangle is the metasimilarity matrix. (Right) H matrix obtained from the K-means partition and the cutting of the tree into four components.

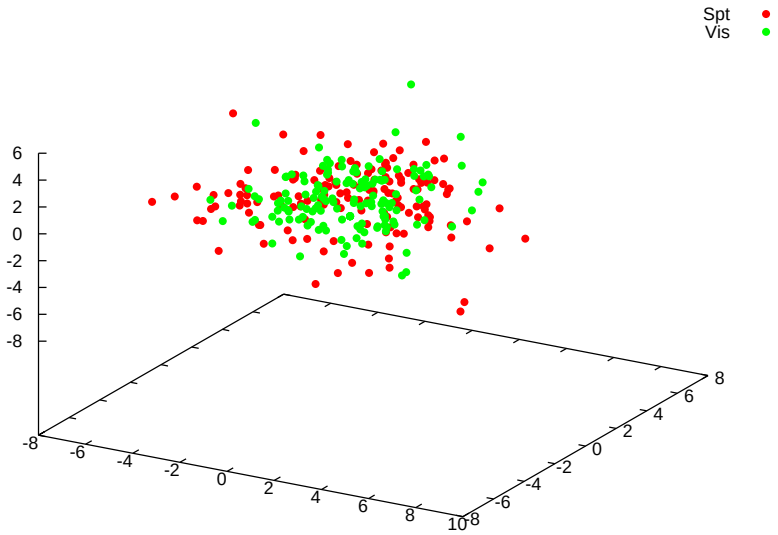


Figure 1.27: **PCA space.** - The on-going activity (red) and the visually evoked activity (green) data sets are projected on the PCAs of the on-going activity data set.

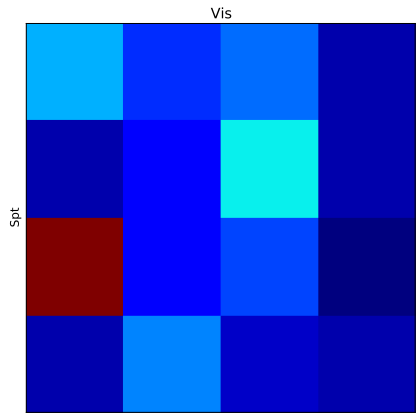


Figure 1.28: **H matrix for Spt and Vis.** - The on-going activity (red) and the visually evoked activity (green) data sets are projected on the PCAs of the on-going activity data set.

biggest cluster of Spt150 with 6 clusters and, during visual stimulation, neuronal dynamics gathers in this central cloud. Cells in this cluster are characterized by a Gaussian distribution of the membrane potential and we suggest that it defines an operating state of the network in which inputs are processed in a fast and efficient manner. Theoretical arguments explaining how such an asynchronous irregular state can be sustained in the network will be given in the next chapter. In the pink cluster, cells have a clear bistable behaviour and transitions between up and down states are correlated with the EEG suggesting a coherent low frequency oscillation at the network level, similar to that observed during slow wave sleep. In the yellow cluster, cells have only very few large deviations from the mean membrane potential suggesting discharge in a synfire mode. An estimation of the number of synchronous inputs generating the synaptic events may be computed from the measurement of the size of these events. Other classes includes cells with micro up or micro down states.

Morphological reconstruction of the neurons could determine whether some of the classes we obtained characterizes a specific cellular type or a cell can access any of the classes. It has been shown, in xylazine-ketamine preparations, that the up state of bistable cell share many similarities with the irregular activity of the awake state [74] so that the same cell could be in the red cluster or the pink cluster depending on the global state of the brain monitored by anesthesia. It would also be interesting to check if classes of neuronal dynamics are related to the states obtained after compression of long EEG recordings and if a cell jumps from a cluster to another during an EEG state transition.

The classes defined above can also be used to investigate how the functional properties of a cell depends on the state of the network in which it is embedded. In a work realized at the UNIC by Nicolas Benech on 118 cells of the data set we studied, it was shown that most of bistable cells have a complex receptive field and that their latency is longer than when cells have monomodal distribution characterizing the operating state. It was shown, in a xylazine-ketamine preparation, that the response to visual stimulation is enhanced when the stimulus is triggered during the up state [75] so that up states may be a cellular analog of attentional facilitation [76].

Chapter 2

The computational units of the brain: neuron and column models.

2.1 Introduction.

In the 19^{th} century, innovative staining techniques of the nervous system by pionneers of histology, like Gerlach and Golgi, were used to support the reticular theory of cerebrospinal organization. The brain tissue was considered to be a continuous net of nerve fibers with holistic properties, being an exception to the cell theory. At the end of the 19^{th}, Ramon y Cajal improved Golgi staining techniques and contributed to the opposite theory: the neuron doctrine. The reticular theory was also in contradiction with the localization of function in the brain like the Broca area dedicated to speech production and discovered in 1861. Sherrigton also supported the neuron theory and named synapse the connexion point between two neurons. The neuron is composed of a dendritic arbor on which presynaptic neurons make contact at dendrites, a cell body, also called soma and an axon. If inputs incoming to the cell body are sufficient, a spike is initiatiated at the axon hillock and propagates through the axon. These parts are illustrated on fig 2.1 for a generic neuron but many cell types with their specific morphology are found in the cortex. This all-or-none behaviour was used to design simplified models capturing the computational properties of the neuron, that is the way inputs are combined before deciding whether to spike or not. In a simple example of such artificial neurons, originally proposed by McCullough and Pitts [77], a weighted sum of the inputs is passed through a sigmoid transfer function. Having interesting computational properties, like any boolean function can be implemented by a network of such units, this artificial neural network, sometimes with different transfer functions and additional learning dynamics on the weights, were a key element in the development of

cybernetics and more specifically connectionism. Beside such formal approach, the understanding of the biophysical mechanisms responsible for spike generation and propagation resulted in more realistic models of the neuron dynamics which will be presented in this chapter.

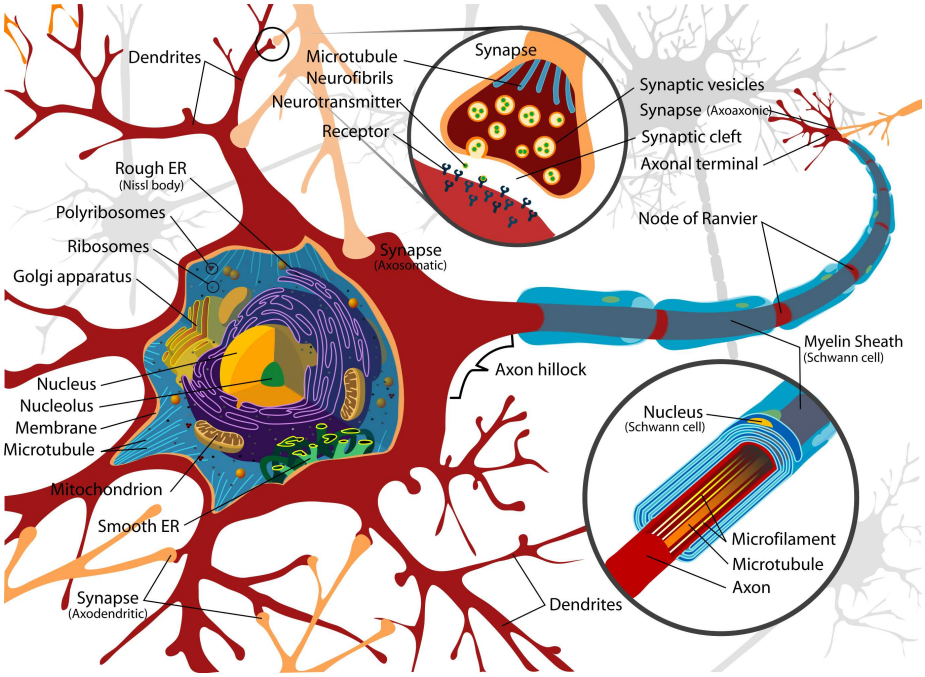


Figure 2.1: **Diagram of a neuron with myalinated axon.** - The dendrites, the soma, the axon and additional structures are indicated (Adapted from http://en.wikipedia.org/wiki/Neuron).

The modular theory of the brain and observations in histology led to the discovery of another set of computational unit at a mesoscale: the cortical columns. Cortical columns have been defined on anatomical ground with minicolumns of $50\mu m$ width containing around a hundred of cells being the result of cell migration during development [78]. The macrocolumn is defined on a functional ground, as described in the introduction, but its anatomical subtrates are estimated of around $300\mu m$ for the visual cortex of the macaque monkey in [79]. A typical macrocolumn contains few thousands of cells and the detailed model of a cortical column of the rat somatosensory cortex in the Blue Brain project contains 10000 neurons of 200 possible types in a space of $500\mu m$ width and $1.5mm$ depth [80]. The column gathers cells coding for the same feature of the inputs so that a feature is reflected in the activity of a population of cells rather than in a single cell spike train. This redundancy in the vertical direction of the cortex makes the code robust to perturbations of the dynamics like synaptic transmission failure or intrinsic fluctuations in cortical dynamics. First

evidences for such columnar organization were found in the somatosensory cortex of the cat by Mounstcastle [81] in the 50s and few years later in Hubel and Wiesel work on the primary visual cortex of the cat [82].

Depending on the animal and the area considered, the neuronal computations can be understood at the single cell level or at the column level, it is thus necessary to analyze models of these computational units with tools from the theory of dynamical systems which will be presented in this chapter. Biologically realistic models of the neuron have several variables (4 in the classical Hodgkin-Huxley model) often following a non-linear evolution equation making their analysis a difficult task. Reduced low dimensional models capture the essential features of the dynamics taking advantage of linearly related variables in the Fitz-Hugh Nagumo model or caricaturating the spike by an instantaneous reset after the membrane crosses a threshold. Models of a cortical column, with their huge state space, can also be reduced by considering the mean field approximation of the network. In this chapter, after presenting the neuron and the cortical column, we give a short introduction to dynamical systems and then apply such methods to models of the computational units of the nervous system.

2.2 Dynamical systems.

A neuron and a column can both exhibit complex dynamics and the theory of dynamical systems is of great use to understand it. The main concepts and some examples are briefly summarized bellow and a full presentation of the bifurcation theory can be found in [83], [84] and [85].

2.2.1 Invariant sets, stability.

A dynamical system consists of the triple T,X,ϕ^t, where T is a time set, X is a state space and ϕ^t is a family of evolution operators parametrized by $t \in T$ and satisfying the following properties: $\phi^0 = id$ and $\phi^{t+s} = \phi^t \circ \phi^s$. Dynamical systems are studied through the orbits they produce and an orbit starting at x_0 is the ordered subset of the state space X, $Or(x_0) = \{x \in X : x = \phi^t x_0$ for all $t \in T$ such that $\phi^t x_0$ is defined$\}$. The phase portrait result from the partitioning of the state space into orbits. Particularly simple orbits consist of fixed points and limit cycles. A point $x_0 \in X$ is a fixed point if $\phi^t x_0 = x_0$ for all $t \in T$. A cycle L_0 is a periodic orbit such that each point $x_0 \in L_0$ satisfies $\phi^{t+T_0} x_0 = \phi^t x_0$ with some $T_0 > 0$, for all $t \in T$. A cycle with no other cycle in the neighborhood is called a limit cycle. Fixed point and limit cycle are two examples of invariant sets that is a subset $S \in X$ such that for each point $x_0 \in S$, $\phi^t x_0 \in S$ for all $t \in T$. Invariant sets more complex than fixed points and cycles are related to chaotic dynamics. An important property of an invariant set is its stability because it determines if nearby orbits will be attracted to this set and then if the invariant will be observable in the dynamics of the system. An invariant set S_0 is Lyapunov stable if for any sufficiently small neighborhood $U \subset S_0$ there exists a neighborhood $V \subset S_0$ such that $\phi^t x \in U$ for all $x \in V$ and all $t > 0$. An

invariant set S_0 is asymptotically stable if there exists a neighborhood $U_0 \subset S_0$ such that $\phi^t \to S_0$ for all $x \in U_0$ as $t \to \infty$. An invariant set is stable if it is both Lyapunov stable and asymptotically stable. A dynamical system can be defined from a system of differential equations, $\dot{x} = f(x)$, where the orbits are the solutions of the system. The fixed point x_0 is stable if the eigenvalues, λ_i, of the Jacobian matrix, $A = f_x(x_0)$, of the system at the fixed point all have negative real part, $Re\lambda_i < 0$. Asymptotic stability of a fixed point x_0 can also be demonstrated if there exists a Lyapunov function for the system near x_0, that is a continuous function defined on a neighborhood of x_0, minimum in x_0 and strictly decreasing on its domain of definition. In some cases, the vector field f defining the dynamical system can be derived from a potential V such that $f = \nabla V$. Fixed points of the system will be critical points of the potential and it will be stable if it is a local minimum.

2.2.2 Bifurcations and normal forms

The phase portrait is a good description of a dynamical system and it can be used to compare different dynamical systems. Thus, two phase portraits are topologically equivalent if there exists a homeomorphism mapping the orbits of one onto the orbits of the other, preserving the direction of time. Fixed points can then be classified into stable node, stable focus, unstable node, unstable focus or unstable saddle (see fig 2.2). If there is no eigenvalue of the Jacobian on the imaginary axis, the fixed point is hyperbolic. When a parameter variation leads to a topologically nonequivalent phase portrait, a bifurcation occurs. The codimension of the bifurcation is the number of independant conditions determining the bifurcation. Informations about these changes are gathered in a bifurcation diagram which represent the phase portrait for various parameter values. An equivalence relation between dynamical systems have to consider a mapping from the parameter space of the first to the parameter space of the second. If we take $\dot{x} = f(x, \alpha)$, $x \in \mathbb{R}^n$, $\alpha \in \mathbb{R}^m$ and $\dot{y} = f(y, \beta)$, $y \in \mathbb{R}^n$, $\beta \in \mathbb{R}^m$, those two dynamical systems are topologically equivalent if there exists a homeomorphism of the parameter space $p : \mathbb{R}^m \to \mathbb{R}^m$, $\beta = p(\alpha)$ and there is a parameter dependent homeomorphism of the phase space h_α: $\mathbb{R}^n \to \mathbb{R}^n$, $y = h_\alpha$, mapping orbits of the first system at parameter values α onto orbits of the second system at parameter values $\beta = p(\alpha)$, preserving the direction of time. The two systems are locally topologically equivalent near the origin if

- there exists a mapping $(x, \alpha) \to (h_\alpha(x), p(\alpha))$ defined in a small neighborhood of $(x, \alpha) = (0, 0)$,
- p is an homeomorphism defined in a small neighborhood of $\alpha = 0$,
- $p(0) = 0$,
- h_α is a parameter dependent homeomorphism defined in a small neighborhood U_α of $x = 0$, with $h_0(0) = 0$, mapping the orbits of the first system in U_α onto the orbits of the second system in $h_\alpha(U_\alpha)$.

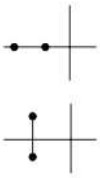
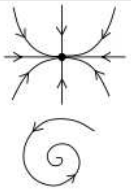
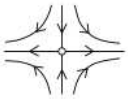
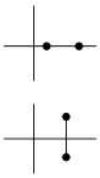
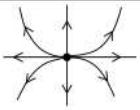
(n_+, n_-)	Eigenvalues	Phase portrait	Stability
(0, 2)		 node focus	stable
(1, 1)		 saddle	unstable
(2, 0)		 node  focus	unstable

Figure 2.2: **Classification of fixed points for a 2D system.** - Depending on the sign of eigenvalues of the Jacobian, a fixed point can be: stable node, stable focus, unstable node, unstable focus or unstable saddle (Adapted from [83]).

A dynamical system defined by a polynomial with a bifurcation of the origin may be topologically equivalent to a simpler polynomial, that is a polynomial of lower degree or taking the symmetry of the system into account. The simplest of such systems is called a topological normal form for this bifurcation if any generic system [1] with an equilibrium satisfying the same bifurcation condition is locally topologically equivalent near the origin to this system for some values of the coefficient of the polynomial.

2.2.3 Examples of bifurcations.

Fold bifurcation.

The most simple bifurcations are related to the loss of hyperbolicity of a fixed point. The fold (also called saddle-node) bifurcation occurs when the Jacobian matrix at a fixed point has a zero eigenvalue. The simple dynamical system $\dot{x} = x^2 + \alpha$ has a nonhyperbolic equilibrium $x_0 = 0$ with $\lambda = f_x(0,0) = 0$ when $\alpha = 0$. When $\alpha < 0$, there are two equilibria $x_\pm(\alpha) = \pm\sqrt{\alpha}$, with x_+ unstable and x_- stable, and when α crosses zero from negative to positive values, the two equilibria collide and disappear so that there is no equilibrium anymore when $\alpha > 0$ as shown on the bifurcation diagram on fig 2.3. It can be shown that any system with higher order terms is locally topologically equivalent to the previous system. Furthermore, any generic dynamical system having a fold bifurcation of the equilibrium $x_0 = 0$ at $\alpha = 0$ is locally topologically equivalent to $\dot{y} = y^2 + \beta$, which is then a normal form for the fold bifurcation. The conditions for the bifurcation to be generic are the non degeneracy condition, $\frac{1}{2} f_{xx}(0,0)0$, and the transversality condition, $f_\alpha(0,0) \neq 0$.

Hopf bifurcation.

If at some parameter value, eigenvalues are $\pm i\omega$, a Hopf bifurcation occurs with a limit cycle emerging from a fixed point. This can only occur in at least 2 dimensional systems and the topological normal forms for such bifurcation is:

$$\begin{cases} \frac{dx}{dt} &= \beta x - y + sx(x^2 + y^2 - \alpha) \\ \frac{dy}{dt} &= x + \beta y + sy(x^2 + y^2 - \alpha) \end{cases}$$

with $s = \pm 1$ depending on the Lyapunov coefficient of the original system. If $s = -1$, the fixed point becomes repelling at $\alpha = 0$ and the activity follows the branch of the stable periodic orbit, this is the supercritical Hopf bifurcation. If $s = 1$, the Hopf bifurcation is subcritical, unstable periodic orbits colliding with the stable node so that it becomes repelling after the bifurcation and the system jumps to the closest stable set (see fig 2.3).

[1] A generic system satisfies:

- A nondegeneracy condition: at least one coefficient of higher order than linear don't vanish at the bifurcation point.
- A transversality condition: derivative of the real part of the eigenvalue with respect to the bifurcation parameter is non zero a the bifurcation point.

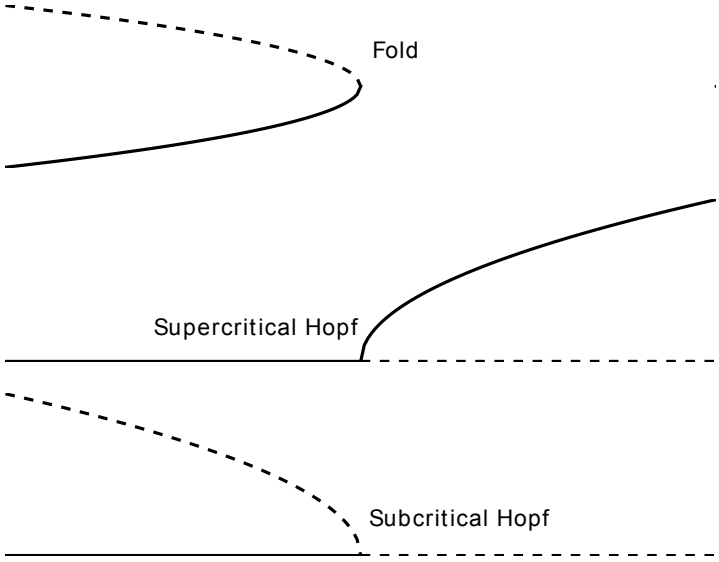


Figure 2.3: **Bifurcation diagrams** - (Up) Fold bifurcation. (Midle) Supercritical Hopf bifurcation. (Down) Subcritical Hopf bifurcation.

Homoclinic orbit.

An orbit starting at x is called homoclinic to the equilibrium x_0 if $\phi^t x \to x_0$ as $t \to \pm\infty$, that is the orbit connects the saddle x_0 to itself. Heteroclinic orbits connect a saddle to another saddle. These phenomena are found for a particular value of the parameter which is an example of global bifurcation.

2.3 Stochastic dynamics.

Often some random behaviour is observed in neuronal data and a noise term is often included in models of neuronal dynamics. Thus to get a good representation of the dynamics, one realization is not enough and it is necessary to consider averages and distributions over an ensemble of realizations of the processes. An introduction to stochastic dynamics and random dynamical systems can be found in [86], [87] and [88].

2.3.1 Stochastic processes.

Definition Given a probability space $(\Omega, \mathcal{F}, P)$, with Ω the set of possible outcomes, $\mathcal{F}$ the set of events and P the probabilities attributed to these events, a stochastic process is a collection of random variables $X_t, t \in T$ defined on $(\Omega, \mathcal{F}, P)$. The process is discrete in time if $T = \mathbf{N}$ and continuous in time if $T = \mathbf{R}^+$. The Markov property characterizes minimal memory processes where the present state is sufficient to get the full distribution over the future, this can be expressed by the following conditional probability distributions:

$$p(x_{k+1}, t_{k+1}; ...; x_n, t_n | x_1, t_1; ...; x_k, t_k) = p(x_{k+1}, t_{k+1}; ...; x_n, t_n | x_k, t_k).$$

A Markovian process checks the Chapman-Kolmogorov equation for transition probabilities:

$$p(x_1, t_1 | x_3, t_3) = \int dx_2 p(x_1, t_1 | x_2, t_2) p(x_2, t_2 | x_3, t_3)$$

.

There are several kinds of stochastic processes:

- Continuous processes where the random variables can take all possible values in their range of definition.
- Jump processes where the random variables can take values on a finite set and jumps from one of this value to another.
- Point processes where event take place at random times.

The definition for the continuity of a stochastic process is given by the Lindeberg condition, for every $\epsilon > 0$:

$$lim_{\Delta t \to 0} \int_{|x-z|>\epsilon} dx p(x, t + \Delta t | z, t) = 0$$

.

The Brownian motion. A simple example of stochastic process checking the Markov property is the Brownian motion [2] $B(t), t \in T$ characterized as follows:

- B(0)=0.
- For all $t_1 < t_2 < ... < t_n$, increments $B(t_2) - B(t_1), ..., B(t_n) - B(t_{n-1})$ are random independent variables.
- For $0 < s < t$, the random variable $B(t) - B(s)$ follows a normal distribution $\mathcal{N}(0, t - s)$
- The transition probabilities of $B(t)$ checks the Lindeberg continuity condition.

The probability distribution of B(t) follows a Gaussian law with mean $\mathbf{E}[B(t)] = 0$ and $\mathbf{E}[B(t)^2] = t$, it is then a solution of the following diffusion equation:

$$\frac{\partial p(x,t)}{\partial t} = \frac{1}{2}\frac{\partial^2 p(x,t)}{\partial x^2}, p(x,0) = \delta(x)$$

having as solution $p(x,t) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}}$.

The increments $\Delta B(t_i)$ for a process can be generated by the Box-Mueller algorithm:

- Pick up U and V, random numbers independent and from a uniform law in $[0, 1]$.
- Compute $X = \sqrt{-2ln(U)}cos(2\pi V)$ and $Y = \sqrt{-2ln(U)}sin(2\pi V)$.

then X and Y are independent random variables with $\mathcal{N}(0, 1)$ as probability density function.

The Poisson process A Poisson process is a counting process, $(N_t)_{t\geq 0}$ a family of random variables indexed by $t \in \mathbf{R}$ and taking values in $\mathbf{N}$ with independent successive increments and the probability distribution of (N_t) only depending on the length of the time intervals. For such a process $(N_t)_{t\geq 0}$ follows a Poisson distribution : $P(N_{t+\tau} - N_t = k) = \frac{e^{\lambda\tau}(\lambda\tau)^k}{k!}$ with λ the rate of the process, times between to events follows an exponential distribution $f(T) = \lambda e^{-\lambda T}$. The average inter events interval is $\mathbf{E}[T] = \frac{1}{\lambda}$ and the variance is $\mathbf{E}[(T - \mathbf{E}[T])^2 = \frac{1}{\lambda^2}$.

Ito processes Ito processes can be written in integral form:

$$x(t) = x(0) + \int_0^t a(x,s)ds + \int_0^t b(x,s)dB(s)$$

[2] It was named after the botanist Robert Brown who observed random trajectories of pollen particles in water and it is sometimes called a Wiener process after the mathematician Norbert Wiener who provided a formalization for it.

or equivalently as a stochastic differential equation [3]:

$$dx(t) = a(x,t)dt + b(x,t)dB(t).$$

Where $a(x,t)$ stands for the deterministic part of the dynamics and $b(x,t)$ for the stochastic part, the noise term is additive if b doesn't depend on x and is multiplicative otherwise. For a multiplicative noise, the integral formula can be interpreted with Ito's definition

$$\int_0^t b(x,s)dB(s) = lim_{n\to\infty} \sum_{i=1}^{n} b(x(t_{i-1}))(B(t_i) - B(t_{i-1}))$$

or with Stratanovich's definition

$$\int_0^t b(x,s)dB(s) = lim_{n\to\infty} \sum_{i=1}^{n} b(\frac{x(t_{i-1}) + x(t_i)}{2}, t_{i-1})(B(t_i) - B(t_{i-1}))$$

but for additive noise the two formulations are equivalent.

2.3.2 Stochastic calculus.

Ito's formula and Forward Kolmogorov equation In stochastic calculus, the chain rule for derivation must be modified because terms of order $(dB(t))^2$ are of same order as dt. The derivative of $y = f(x)$ is then at first order in dt:

$$dy(t) = (\frac{\partial f(x)}{\partial x}a(x,t) + \frac{1}{2}\frac{\partial^2 f(x)}{\partial x^2}b^2(x,t))dt + \frac{\partial f(x)}{\partial x}b(x,t)dB(t).$$

Applying this formula to $\mathbf{E}[f(x)] = \int f(x,t)p(x,t)dx$, with p the probability distribution of x, gives:

$$\frac{d}{dt}\int f(x,t)p(x,t)dx = \int [\frac{\partial f(x)}{\partial x}a(x,t) + \frac{1}{2}\frac{\partial^2 f(x)}{\partial x^2}b^2(x,t)]p(x,t)dx$$

which can be integrated by parts:

$$\frac{d}{dt}\int f(x,t)p(x,t)dx = \int [-\frac{\partial p(x,t)a(x,t)}{\partial x} + \frac{1}{2}\frac{\partial^2 p(x,t)b^2(x,t)}{\partial x^2}]f(x,t)dx.$$

As this is checked for any f, it gives the forward Kolmogorov or Fokker-Planck equation [4]:

$$\frac{dp(x,t)}{dt} = -\frac{\partial p(x,t)a(x,t)}{\partial x} + \frac{1}{2}\frac{\partial^2 p(x,t)b^2(x,t)}{\partial x^2}.$$

It can be generalized to N-dimensional processes so that:

$$d\mathbf{x}(t) = \mathbf{a}(\mathbf{x},t)dt + \mathbf{b}(\mathbf{x},t)d\mathbf{B}(t)$$

[3] called the Langevin equation.

[4] It can also be derived as the Kramers-Moyal expansion of the Chapman-Kolgomogorov equation truncated at order 2, see [87].

have the following Fokker-Planck equation:

$$\frac{\partial p(\mathbf{x},t)}{\partial t} = \sum_{1\leq i\leq N} \frac{\partial a_i((x),t)p(\mathbf{x},t)}{\partial x} + \sum_{1\leq i\leq N}\sum_{1\leq j\leq N} \frac{\partial^2 \mathbf{b}(\mathbf{x},t)\mathbf{b}^T(\mathbf{x},t)p(\mathbf{x},t)}{\partial x_i \partial x_j}.$$

By introducing probability currents, for $1 \leq i \leq N$:

$$J_i = a_i(\mathbf{x},t)p(\mathbf{x},t) - \frac{1}{2}\sum_{1\leq j\leq N} \frac{\partial \mathbf{b}_{ij}(\mathbf{x},t)p(\mathbf{x},t)}{\partial x_j},$$

the Fokker-Planck equation can be written as a conservation law for the probability density:

$$\frac{\partial p(\mathbf{x},t)}{\partial t} + \sum_{1\leq j\leq N} \frac{\partial J_i(\mathbf{x},t)}{\partial x_j} = 0$$

Boundary conditions must be added to be able to solve this equation, it is commonly taken as $lim_{|\mathbf{x}|\to\pm\infty} p(\mathbf{x},t) = 0$ but absorbing or reflecting barriers may be specified.

2.4 Numerical integration and analysis.

We now present some numerical integration schemes for ordinary differential equations with a special focus on explicit methods. In an explicit method x_{n+1} only depends on previous values x_k whereas in implicit methods it alsodepends on itself.

2.4.1 Integration of deterministic systems.

Euler scheme. For a one dimensional dynamical system defined by the following differential equation:

$$\frac{dx(t)}{dt} = f(t, x(t))$$

a trajectory starting at $x(t_0) = x_0$ can be integrated with a time step h by the Euler method, for $n > 0$:

$$x_{n+1} = x_n + hf(t_n, x_n)$$

which is just the approximation obtained by considering the first terms of the Taylor expansion. The higher order approximation is

$$x_{n+1} = x_n + hf(x_n) + h^2(\frac{\partial f}{\partial t}(t_n, x_n) + \frac{\partial f}{\partial x}(t_n, x_n)f(t_n, x_n))$$

so that the error between the numerical solution and the exact solution scales as h^2 thus being of order 1.

Heun scheme The Heun method is another one step integration scheme:

$$x_{n+1} = x_n + \frac{h}{2}(f(t_n, x_n) + f(t_n, u_n + hf(t_n, x_n)))$$

but it is of order 2 so that the error scales as h^3.

Runge Kutta scheme When f is non-linear, errors from the Euler method can be reduced by employing more sophisticated methods like the Runge Kutta methods which includes multiple steps. In the fourth order method, two intermediate points are introduced, so that:

$$x_{n+1} = x_n + \frac{h}{6}\sum_{i=1}^{4} b_i k_i$$

with $b_1 = 1$, $b_2 = 2$, $b_3 = 2$, $b_4 = 1$ and

$$k_1 = f(t_n, y_n)$$

$$k_2 = f(t_n + \frac{h}{2}, x_n + \frac{h}{2}k_1)$$

$$k_3 = f(t_n + \frac{h}{2}, x_n + \frac{h}{2}k_2)$$

$$k_4 = f(t_{n+1}, x_n + \frac{h}{2}k_3)$$

A s stage Runge Kutta method cannot be of order higher than s and this method can be extrapolated to an arbitrary number f stages increasing the accuracy. In the Gill's method, coefficients are a bit modified: $b_1 =, b_2 = 2 - \sqrt{2}, b_3 = 2 + \sqrt{2}, b_4 = 1$ and

$$k_1 = f(t_n, x_n)$$

$$k_2 = f(t_n + \frac{h}{2}, x_n + \frac{h}{2}k_1)$$

$$k_3 = f(t_n + \frac{h}{2}, x_n + \frac{h}{2}(-1 + \sqrt{2})k_1 + h(1 - \frac{\sqrt{2}}{2})k_2)$$

$$k_4 = f(t_n + h, x_n - \frac{h\sqrt{2}}{2}k_1 + (1 + \frac{\sqrt{2}}{2})k_3).$$

These methods can also be improved to implement time step adaptivity.

2.4.2 Integration of stochastic systems

Numerical integration scheme can also be used to integrate stochastic dynamics.

Euler-Maruyama method. For a stochastic differential equations, like with additive noise $dx(t) = f(t,x)h + \sigma(x)dB(t)$, the Euler method can be modified by introducing the stochastic term: $x_{n+1} = y_n + hf(t_n, x_n) + \sigma(x_n)\sqrt{h}(B(t_{n+1}) - B(t_n)$ where $\Delta(B_t) = B(t_{n+1}) - B(t_n)$ are independent and identically distributed random variables of mean 0 and variance 1.

Milstein method Milstein's scheme is as follows:

$$\begin{aligned} x_{n+1} &= y_n + hf(t_n, x_n) + \sigma(x_n)\sqrt{h}(B(t_{n+1}) - B(t_n)) \\ &\quad + h\sigma(x_n)\sigma'(x_n)((B(t_{n+1}) - B(t_n))^2 - 1) \end{aligned}$$

2.4.3 Analysis of dynamical systems.

Spectral analysis. The Fourrier spectrum of the raw stochastic signal is not the best tool for the analysis of a stochastic signal because it will depend on the specific realization of the noise term. The autocorrelation is a better solution for describing properties of the signal. For a signal $y(t)$, the Fourrier transform is $\hat{y}(\omega) = \int_{-\infty}^{\infty} y(t)e^{-i\omega t}dt$ and the autocorrelation is $< y(t+\tau)y^*(t) >= \int\int_{-\infty}^{\infty} P(x_1, t+\tau, x_2, t)dx_1dx_2$. The Wiener-Kintchine theorem then relates the spectral density $S(\omega) =< \hat{y}(\omega)\hat{y} * (0) >$ to the autocorrelation $< y(\tau)y * (0) >$ by $S(\omega) = \int_{-\infty}^{\infty} e^{-i\omega\tau} < y(\tau)y * (0) > d\tau$.

Lyapunov exponent. Dynamical systems are said to be chaotic if their trajectories diverge exponentially. For trajectories separated by δx_0 at initial time, the difference grows as $\delta x(t) = \delta x_0 e^{\lambda t}$ and the growth coefficient λ is a good indicator of chaotic systems when it is positive, it is called the Lyapunov exponent of the system. For multi-dimensional sytems, there are several directions along which coefficients can be contracting ($\lambda < 0$) or expanding ($\lambda > 0$). If there exist an invariant set, like a limit cycle, $\lambda = 0$ along this set. For smooth dynamical systems, chaotic trajectories shows up only when the dimension is at least three. The Lyapunov spectrum (the set of Lyapunov exponents) is usually calculated by following the dynamics of along the Jacobian and then calculating the expansion and contraction rates. For a system

$$\dot{x} = f(x),$$

the equivalent linear system is

$$\dot{u} = Df(x)u.$$

To avoid accumulation of the dynamics of perturbations along the direction corresponding to the maximal Lyapunov exponent, a Gram-Schmidt orthonormalization procedure is usually adopted [89] transforming a set of vectors $(u_1, ..., u_n)$ into a orthonormal basis of $\mathbf{R}^n$ $(v1, ..., v_n)$:

$$w_1 = u_1, v_1 = \frac{w_1}{\|w_1\|}$$

$$w_2 = u_2 - (u_2.v_1)v_1, \frac{w_2}{\|w_2\|}$$
$$\ldots$$
$$w_n = u_n - \sum_{k=1}^{n-1}(u_n.v_k)v_k, \frac{w_n}{\|w_n\|}$$

This orthonormalization procedure is done every T, K times, while equations for x and u are integrated. The i^{th} Lyapunov exponent is computed as $\lambda_i =< ln\|w_i\| >$ with $< . >$ is the average over iterations.

2.5 Neuron

Neurons have a huge diversity in their structure and their dynamic properties. The Neurolex initiative [5] tries to build a common language for their classification following the Petilla classification for GABAergic neurons [90]. We first describe the diversity of cells encountered in the brain and then we show how simple models can account for essential features the dynamics of the membrane potential in spite of the cellular diversity.

2.5.1 Diversity of the cells.

Excitatory and inhibitory cells. The major classification of cells is on their influence to other neurons which is mediated for chemical synapses by neurotransmitters flowing at the synaptic cleft which is the $20nm$ space between axons terminals and dendritic buttons (see fig 2.1). Neuronal interactions depends on the receptor type, AMPA [6] and NMDA [7] synapses are excitatory whereas GABAa and GABAb [8] synapses are usually considered as inhibitory although these synapses have excitatory effect in early developmental stages. Neurons can also interact through electrical synapses, also called, where the signal can be transmitted faster than for chemical and often bidirectionally via physical contact between the two neurons. The effect of such synapses can be depolarizing or hyperpolarizing depending on the presynaptic activity.

Structure. Another way to classify neurons is on their structure. The most common and biggest neuron in the cortex is the excitatory pyramidal cell which has a triangular soma, a dense dendritic tree with apical and basal parts and a long myelinated axon. Another class of excitatory cells are spiny stellate cells having a symmetric star shape with localized axon. The majority of inhibitory interneurons in layer IV are basket cells. Another class of GABAergic inhibitory interneurons, the chandelier cell, is named after the shape of its axon terminals and the Purkinje cell, also GABAergic but located in the cerebellar cortex, is famous for its beautiful planar dendritic arbor.

[5] Available at http://www.neurolex.org.
[6] AMPA stands for α-amino-3-hydroxy-5-methylisoazol-4-propionate
[7] NMDA stands for N-methyl-D-aspartic acid
[8] GABA stands for gamma-aminobutyric acid.

Channels. The membrane of a neuron includes many voltage gated channels letting ions flow inside or outside the cell depending on the membrane voltage. These channels can also be used to classify neurons. Potassium and sodium channels are present for all neuron types and explain spike generation in the Hodgkin-Huxley model. Other channels are involved in more specific processes like the T-type voltage-gated calcium channel, responsible for the tonic bursting of thalamic cells.

Firing pattern. Neurons can also be classified based on their firing properties in response to a step of input current [91]. Inhibitory interneurons are often fast spiking cells with constant and short interspike interval (ISI). Excitatory cells have different patterns like non-adapting regular spiking with constant ISI but longer than for fast spiking cells. For adapting regular spiking cells, ISI decreases during the response. Intrinsic bursting cells fire with few spikes very close together at the response onset. The Petilla classification includes more complex firing patterns like stuttering, irregular or accelerating. For some neurons, the study of the after potential hyperpolarization can also be helpful for classification.

2.5.2 Dynamic processes.

Models of the dynamics of the membrane potential should include a spike generation mechanism and also describe the synaptic interaction.

Spike generation. The generation of spikes is attributed to two ionic concentrations: K^+ and Na^+. At rest, potassium ions are in excess outside the cell and sodium ions are in excess inside the cell. When the concentrations of these ions inside and outside the cell are balanced to reach equilibrium, the corresponding difference of potential between the inside and the outside is given by the Nernst potential defined as

$$E_{Nernst} = \frac{RT}{z\mathcal{F}} ln \frac{c_{out}}{c_{in}},$$

with

$$R = 6.02.10^{23} mol^{-1}$$

the Avogadro constant ,

$$\mathcal{F} = 9.6510^{-4} C.mol^{-1}$$

the Faraday constant, T the temperature, z the number of charges carried by the ion and c_{in},c_{out} the ion concentrations inside and outside the cell. When the membrane potential deviates from this value, a ionic current is generated proportional to the deviation $(V - E_{Nernst})$ (with $E_K = -77mV$, $E_{Na} = 50mV$). Moreover, the coefficient of proportionality of this current is constant for a passive channel like the leak current $I_L = g_L(V - E_L)$ (with $E_L = -65mV$) but

it depends on the voltage for active channels like the Na and K channels. Each ion flux is conditioned on the opening or closing of active gates. Each gate can be open or closed with transition probability from open to close $\alpha(V)$ and from closed to open $\beta(V)$. The fraction of open channels follows the dynamics:

$$\frac{dm}{dt} = \alpha(V)(1-m) - \beta(V)m$$

or equivalently:

$$\frac{dm}{dt} = \frac{(m_\infty - m)}{\tau(V)}$$

with $m_\infty = \frac{\alpha(V)}{\alpha(V)+\beta(V)}$ and $\tau(V) = \frac{1}{\alpha(V)+\beta(V)}$. In the Hodgkin-Huxley model, 3 channel variables are considered: one for the fraction of activation of the potassium gate n, one for the fraction of activation of the sodium gate m and one for the fraction of inactivation of the sodium gate $1-h$. The current corresponding to these ionic transports are

$$I_K = g_{\bar{K}} n^4 (V - E_K)$$

and

$$I_N a = g_{\bar{N}a} h m^3 (V - E_{Na})$$

with $\bar{g}_K$ and $\bar{g}_{Na}$ the maximal conductances. When the cell is slightly depolarized Na channels open and flow inside the cell, while V is increasing until the driving current proportional to $(V - E_{Na})$ becomes small and Na gets inactivated. Then K currents activate and the potassium driving force $(V - E_K)$ is strong so that the membrane potential decreases and return to its resting value.

Synapse dynamics. An action potential propagates along the axon and when it reaches the synaptic terminals, neurotransmitters are released in the synaptic cleft and postsynaptic events are triggered in the postsynaptic neuron. The corresponding current is

$$I_{syn} = g_{syn}(t)(V(t) - E_{syn})$$

with the synaptic conductances $g_{syn}(t) = \bar{g}_{syn} s(t)$ generated from the incoming spikes by the following dynamics:

$$\tau_{syn} \frac{ds}{dt} = -s + \tau_{syn} \sum_k \delta(t - t_k)$$

where k runs over all presynaptic spikes and the Dirac impulse defined as $\delta(x) = 0$ for $x = 0$ and 0 elsewhere. The solution for s is a sum of exponential $s(t) = \sum_k e^{-\frac{(t-t_k)}{\tau_{syn}}}$. When the dynamics for s is of second order:

$$\frac{d^2 s}{dt^2} + \frac{2}{\tau_{syn}} \frac{ds}{dt} + \frac{1}{\tau_{syn}^2} s = \frac{1}{\tau_{syn}^2} \sum_k \delta(t - t_k),$$

the solution is a sum of α-functions:

$$s(t) = \sum_k \frac{t}{\tau_{syn}^2} e^{-\frac{(t-t_k)}{\tau_{syn}}}.$$

The synaptic reversal potential is $E_E = 0mv$ for excitatory synapses and $E_I = -80mV$ for inhibitory synapses. For electrical synapses, the synaptic current are: $I_{syn} = \bar{g}_{syn}(V_{post} - V_{pre})$. Slow dynamics, like depression, facilitation or spike timing dependent plasticity can also be included in a synapse [92] [93].

Membrane equation. The membrane of the neuron can be seen as a capacitive medium and an equivalent cable equation can be written to describe the propagation of an action potential along the passive parts of the cell the Rall model:

$$C\frac{dV}{dt} = -V + \frac{d^2V}{dx^2}.$$

Active currents responsible for the generation of the action potential are included in a space clamped version of the model with the equivalent circuit drawn on fig . Applying the Kirchoff law in this circuit gives the Hodgkin-Huxley equation:

$$C\frac{dV}{dt} = -I_m - I_E - I_I$$

with $I_m = \sum_i g_i(V - E_i)$ $(i = K, Na, L)$ the intrinsic currents and $I_{E,I} = g_{E,I}(V - E_{E,I})$ the synaptic currents.

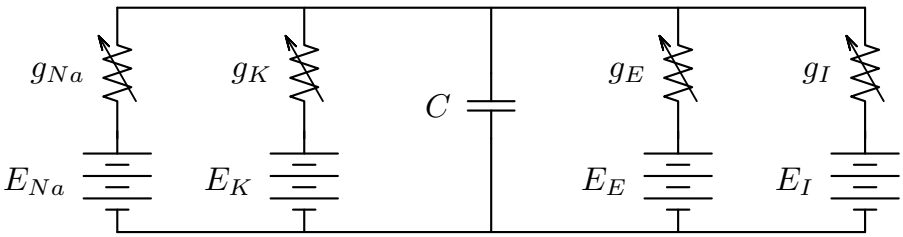


Figure 2.4: **Equivalent circuit for the Hodgkin-Huxley neuron.** - Voltage-gated sodium and potassium channels (g_{Na}, g_K) and synaptic channels and synaptic channels (g_E, g_I). The capacity C scales the membrane time constant.

2.6 FitzHugh Nagumo model of cell excitability

2.6.1 Derivation of the model

The full Hodgkin-Huxley (HH) system of equations describing the dynamics of the membrane potential of a neuron is difficult to study and in th 60's, it was difficult to simulate. Fitz-Hugh and Nagumo thus used the simplified system

which shares many properties with the full system and make a geometric analysis feasible. The system is composed of two variables at different time scales tuned by the parameter ϵ, the limit $\epsilon \to 0$ making it a slow-fast system.

Starting from the HH equations, two approximations reduce the 4-dimensional system to a 2-dimensional system. The full HH system takes sodium, potassium and leak currents into account:

$$\begin{aligned}
C_m \frac{dV}{dt} &= I(t) - \overline{g}_l(V - E_l) - \overline{g}_K n^4 (V - E_K) - \overline{g}_{Na} m^3 h (V - V_{Na}) \\
\tau_n(V) \frac{dn}{dt} &= n_\infty(V) - n(t) \\
\tau_h(V) \frac{dh}{dt} &= h_\infty(V) - h(t) \\
\tau_m(V) \frac{dm}{dt} &= m_\infty(V) - m(t)
\end{aligned}$$

with

$$\begin{aligned}
\tau_{(n,h,m)} &= \frac{1}{\alpha_{(n,h,m)} + \beta_{(n,h,m)}}, \\
(n_\infty, h_\infty, m_\infty) &= \frac{\alpha_{(n,h,m)}}{\alpha_{(n,h,m)} + \beta_{(n,h,m)}}, \\
\alpha_n &= \frac{.01(V+55)}{1 - e^{-.1(V+55)}}, \\
\beta_n &= .125 e^{-.0125(V+65)} \\
\alpha_h &= 0.07 e^{-.05(V+65)}, \\
\beta_h &= \frac{1}{1 + e^{-.1(V+35)}}, \\
\alpha_m &= \frac{.1(V+40)}{1 - e^{-.1(V+40)}}, \\
\beta_m &= 4 e^{.0556(V+65)}.
\end{aligned}$$

Simulations of these dynamics suggest that some approximations leading to a simpler formulation. The dynamics in the middle panel of figure 2.5 shows that variations of m are quasi-instantenous, so that $m(t) \approx m_\infty(V)$ and the model can be reduced to become 3-dimensional. As seen in the bottom panel of figure 2.5 and in the left panel of figure 2.6, n and h are close to the relation $1.1n(t)+h(t) \approx 0.89$, that brings the model to a 2-dimensional simplified system:

$$\begin{aligned}
C_m \frac{dV}{dt} &= I(t) - \overline{g}_l(V - E_l) - \overline{g}_K n^4 (V - E_K) \\
&\quad - \overline{g}_{Na} \overline{m}_\infty(V)(0.89 - 1.1n)(V - V_{Na}) \\
\tau_n(V) \frac{dn}{dt} &= n_\infty(V) - n(t)
\end{aligned}$$

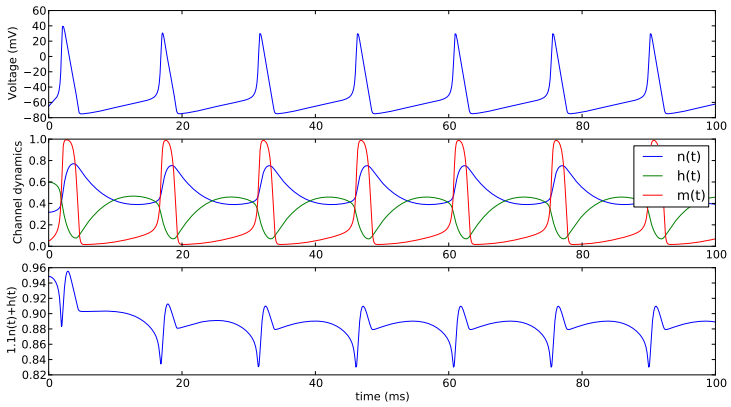


Figure 2.5: Simulation of the Hodgkin-Huxley model. (Up) Dynamics of the membrane potential. (Middle) Dynamics of the activation and inactivation variables of the ionic channels. (Down) 1.1 n(t)+h(t).

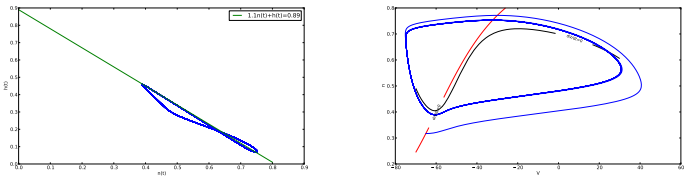


Figure 2.6: (Left) (n,h) dynamics and the line 1.1 n+h=0.89 (Right) (V,n) dynamics and nullclines of the reduced 2D system

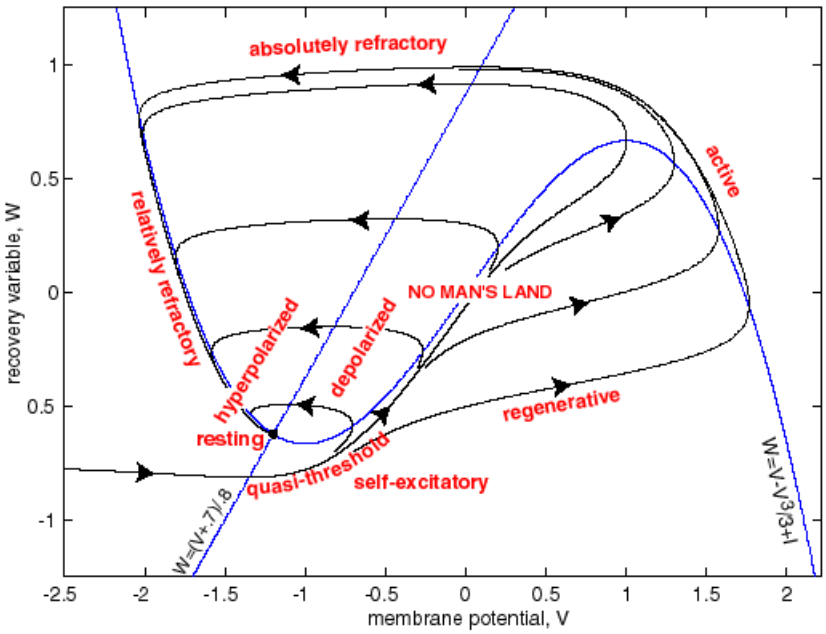


Figure 2.7: Phase portrait of Fitzhugh Nagumo system (Adapted from the Scholarpedia.org article on the FitzHugh-Nagumo model

In the right panel of figure 2.6, the V-nullcline has a N-shape so that it can be approximated by a cubic function and the slow nullcline can be approximated by a straight line leading to the FitzHugh-Nagumo equations:

$$\begin{cases} \epsilon \frac{dx}{dt} &= x - \frac{x^3}{3} + y \\ \frac{dy}{dt} &= -(x - a + by) \end{cases}$$

with $\epsilon > 0$.

2.6.2 Fixed points, saddle-node bifurcation, cusp

Nullclines are

$$\begin{cases} y &= -x + \frac{x^3}{3} \\ y &= \frac{a-x}{b} \end{cases}$$

Fixed points are at the intersection of these lines and are the roots of $\frac{x^3}{3} + (-1 + \frac{1}{b})x - \frac{a}{b}$ which is of the form $x^3 + px + q = 0$ and can be solved using Cardan's method.

Here $p = 3(-1 + \frac{1}{b})$ and $q = -3\frac{a}{b}$, the discriminant is $\Delta = q^2 + \frac{4}{27}p^3$:

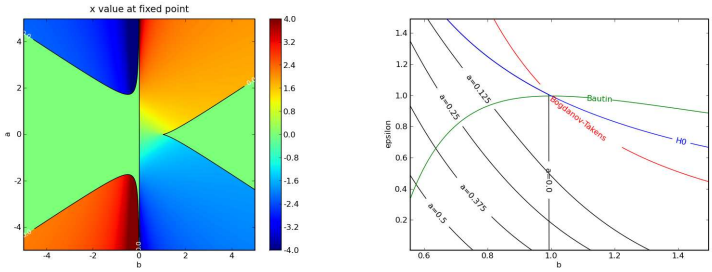


Figure 2.8: (Left) x_0 plotted in color depending on a and b. There are three fixed points in the green areas, the black line is the curve of fold bifurcations and a cusp bifurcation occur at $(1, 0)$.(Right) The red area is where $det(J_{|x_0}) > 0$

- If $\Delta \geq 0$, there is only one real root to the system :

$$x_0 = \sqrt[3]{\frac{-q + \sqrt{\Delta}}{2}} + \sqrt[3]{\frac{-q - \sqrt{\Delta}}{2}}$$

- if $\Delta < 0$ it has 3 solutions($k \in 0, 1, 2$):

$$x_k = 2\sqrt{\frac{-p}{3}} \cos(\frac{1}{3} \arccos(\frac{-q}{2}\sqrt{\frac{27}{-p^3}}) + \frac{2k\pi}{3})$$

A **fold bifurcation** occurs when $det(J_{|x_0}) = \frac{-b(1-x_0^2)+1}{\epsilon} = 0$, it is the line separating the 3 fixed points zone from the one fixed point zone as shown on fig 2.8. When $a = 0$, fixed points are roots of $-\frac{x^3}{3} + (1 - b)x + a$ which is the normal form of a **cusp bifurcation** at $(b = 1, a = 0)$.

2.6.3 Stability of the fixed point when $\Delta > 0$, Hopf and generalized Hopf

Stability matrix near equilibrium x_0:

$$J_{|x_0} = \begin{pmatrix} \frac{1-x_0^2}{\epsilon} & \frac{1}{\epsilon} \\ -1 & -b \end{pmatrix}$$

thus $det(J_{|x_0}) = \frac{-b(1-x_0^2)+1}{\epsilon}$ and $tr(J_{|x_0}) = \frac{1-x_0^2}{\epsilon} - b$.
The characteristic equation is given by:

$$\begin{aligned} P(\lambda) &= \lambda^2 - tr(J_{|x_0})\lambda + det(J_{|x_0}) \\ &= \lambda^2 - (-b + \frac{(1 - x_0^2)}{\epsilon})\lambda + \frac{-b(1 - x_0^2) + 1}{\epsilon} \end{aligned}$$

$$\Delta' = tr(J_{|x_0})^2 - 4det(J_{|x_0}) = (-b + \frac{(1-x_0^2)}{\epsilon})^2 - 4\frac{-b(1-x_0^2)+1}{\epsilon}$$

and eigenvalues of $J_{|x_0}$ are

$$\lambda_\pm = \frac{tr(J_{|x_0}) \pm \sqrt{tr(J_{|x_0})^2 - 4det(J_{|x_0})}}{2}$$

thus damping is $\mu = \frac{1}{2}tr(J_{|x_0})$ and frequency modulation is $\omega = \frac{1}{2}\sqrt{tr(J_{|x_0})^2 - 4det(J_{|x_0})}$.
Andronov-Hopf bifurcation occurs when:

$$\begin{aligned} tr(J_{|x_0}) &= (-b + \frac{(1-x_0^2)}{\epsilon}) = 0 \\ det(J_{|x_0}) &= \frac{-b(1-x_0^2)+1}{\epsilon} > 0 \end{aligned}$$

at the bifurcation point $b = \frac{(1-x_0^2)}{\epsilon}$ and $\lambda_\pm = \pm i\sqrt{-b^2 + \frac{1}{\epsilon}} = \pm i\omega_0$.

To get the normal form, we change coordinates introducing the following vectors:

$q = \begin{pmatrix} -(b+\lambda_+) \\ 1 \end{pmatrix}$ is eigenvector of $J_{|x_0}$ related to λ_+.

$p = \begin{pmatrix} \epsilon(b+\lambda_-) \\ 1 \end{pmatrix}$ is eigenvector of ${}^T J_{|x_0}$ related to λ_-.

The scalar product is $< p, q >= (\hat{p_1}q_1 + \hat{p_2}q_2) = 2\epsilon b(-b + i\omega)$.

We then normalize p taking: $p \to \frac{p}{<p,q>}$ so that: $p = \begin{pmatrix} -\frac{i}{2\omega} \\ 1 - i\frac{b}{\omega} \end{pmatrix}$

We now make the change of coordinates $z =< p, x >$ and with F the non-linear part of the system, the complex variable z is solution of the system:

$$\dot{z} = \lambda_+ z + < p, F(zq, \overline{zq}) >= \lambda z + \sum_{1 \le k+l \le 3} \frac{g_{kl}}{k!l!} z^k \overline{z}$$

$$g(z, \overline{z}) =< p, F(X_0 + zq + \overline{zq}) >$$

$$g(z, \overline{z}) = \frac{i}{2\epsilon\omega}(-\frac{(zq)^3}{3} - \frac{(\overline{zq})^3}{3} - zq(\overline{zq})^2 - (zq)^2\overline{zq} - x_0(zq)^2\overline{zq} - x_0 zq(\overline{zq})^2 - 2x_0 zq\overline{zq})$$

Coefficients of the Taylor expansion are $g_{20} = \frac{-ix_0}{\omega\epsilon^2}((2b^2\epsilon - 1) + 2\epsilon\omega b)$, $g_{11} = \frac{-ix_0}{\omega\epsilon^2}$, $g_{21} = \frac{-\omega + ib}{\omega\epsilon^2}$.

It can be shown that z can be changed to a variable w which after rescaling of the time is solution of the normal form:

$$\dot{w} = (\beta + i)w + l_1 w|w|^2$$

with $\beta = \frac{\mu}{\omega}$ and $l_1 = Re\frac{1}{2\omega}(ig_{20}g_{11} - 2i|g_{11}| - \frac{i}{3}|g_{02}| + \omega g_{21})$ and more simply at the bifurcation :

$$l_1 = \frac{1}{2\omega^2} Re(ig_{20}g_{11} + \omega g_{21})$$

.

$$l_1 = \frac{-\epsilon\omega^2 + 2x_0^2 b}{\epsilon^3 \omega}$$

The sign of the first Lyapunov coefficient characterizes the Hopf bifurcation:

- If $l_1 < 0$, a stable limit cycle emerges at the bifurcation point, it is a **supercritical bifurcation**.
- If $l_1 > 0$, an unstable limit cycle emerges at the bifurcation point, it is a **subcritical bifurcation**.
- **Bautin bifurcation** occurs when $l1 = 0$, that is when $\epsilon = \frac{2b-1}{b^2}$.

There are two possible normal forms for the Bautin bifurcation depending on the sign of the higher order terms (second Lyapunov coefficient l_2):

$$\dot{z} = (\beta_1 + i)z + \beta_2 z|z|^2 \pm z|z|^4$$

with coefficients $\beta_1 = \frac{Re(\lambda)}{Im(\lambda)}$ and $\beta_2 = \sqrt{l_2} l_1$. A polar coordinate transformation, $z = re^{i\phi}$, of the - normal form gives:

$$\left\{ \begin{array}{rcl} \dot{\rho} & = & \rho(\beta_1 + \beta_2\rho^2 - \rho^4), \\ \dot{\phi} & = & 1. \end{array} \right.$$

The only equilibrium is $\rho = 0$ and $\beta_1 + \beta_2\rho^2 - \rho^4 = 0$ may have zero, two or only one solution at the fold of cycles point. There are thus two branches starting from the Bautin point:

- A line H of Hopf bifrcations, $beta_1 = 0$, supercritical for $\beta_2 < 0$ and subcritical for $\beta_2 > 0$.
- A curve $\beta^2 + 4\beta_1 = 0$ with $\beta_2 > 0$, where two limit cycles collide.

Bogdanov-Takens bifurcation occurs when:

$$\left\{ \begin{array}{rcl} tr(J_{|x_0}) & = & (-b + \frac{(1-x_0^2)}{\epsilon}) = 0 \\ det(J_{|x_0}) & = & \frac{-b(1-x_0^2)+1}{\epsilon} = 0 \end{array} \right.$$

that is when $\epsilon = \frac{1}{b^2}$. The normal forms for this bifurcation are:

$$\left\{ \begin{array}{rcl} \dot{\eta_1} & = & \eta_2, \\ \dot{\eta_2} & = & \beta_1 + \beta_2\eta_1 + \eta_1^2 \pm \eta_1\eta_2. \end{array} \right.$$

Moreover, there are three branches passing through the Bogdanov-Takens bifurcation point:

- The line, $\beta_1 = 0$, of Hopf bifurcations.

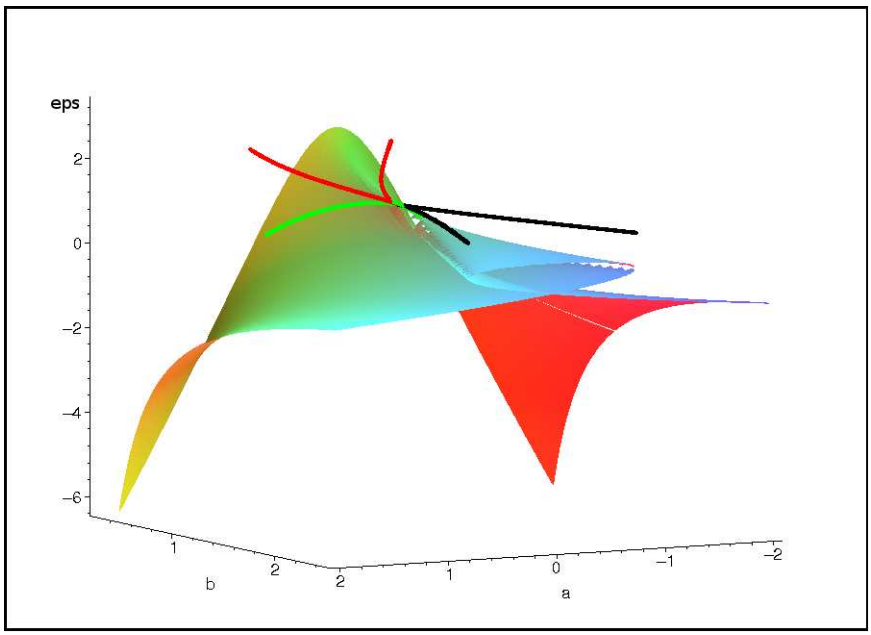


Figure 2.9: (Colored surface) Hopf bifurcation curves ($tr(J_{|x_0}) = 0$) for various values values of a. (Black curve) Fold bifurcation for $\epsilon = 1$, it is the invariant along the ϵ axis. (Red line) Bogdanov-Takens bifurcation curve ($det(J_{|x_0}) = 0$). (Green line) Bautin bifurcation curve ($l_1 = 0$).

- The parabola, $4\beta_1 - \beta_2^2 = 0$, of fold bifurcations.
- The half parabola , $\beta_1 = -\frac{6}{25}\beta_2^2$, of saddle homoclinic bifurcation.

The different types of bifurcation (fold, Hopf, Bautin and Bogdanov-Takens) are represented on the 3D parameter space in fig 2.9. The vertical line of cusp bifurcations,($b = 1, a = 0$), and curves for Bautin and Bogdanov-Takens bifurcations intersect at ($b = 1, a = 0, \epsilon = 1$), this singular situation indicates that this point is a codimension three bifurcation point. It is the organizing center for the dynamics and any behavior of the system is accessible in its neighborhood (except those associated with the slow-fast limit $\epsilon \to 0$).

A normal form for the codimenson three bifurcation point was found along the cusp line in [94]. The system is shown to be topologically equivalent to the following system:

$$\begin{cases} \dot{x_1} &= y_1 - \frac{x_1^3}{3}, \\ \dot{y_1} &= -\frac{x_1^3}{3}. \end{cases}$$

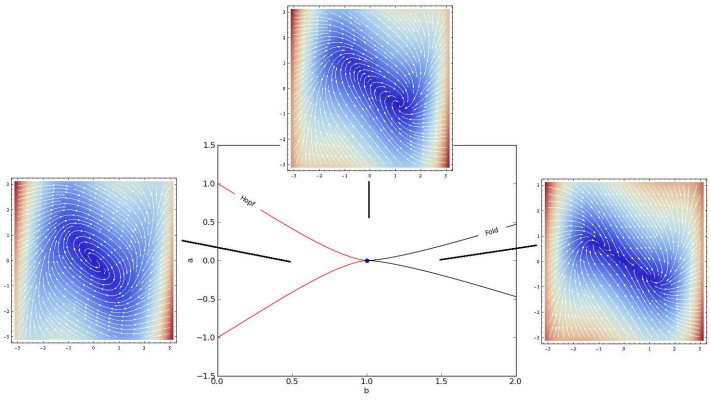


Figure 2.10: Fold and Hopf bifurcation curves starting from the organizing center $\epsilon = 1$ and examples of vector fields.

2.6.4 Dynamics at the organizing center with noisy input.

Bifurcation diagram near $(a = 0, b = 1, \epsilon = 1)$ with example of vector fields (one fixed point, two fixed points and limit cycle) are illustrated on fig 2.10. There is a limit cycle inside the Hopf region, there are 3 fixed points, two of which are stable, inside the fold region and there is a single fixed point elsewhere.

By adding a random component being a Brownian motion, the dynamics writes:

$$\left\{ \begin{array}{rcl} \epsilon dx & = & (x - \frac{x^3}{3} + y)dt + \sigma\sqrt{dt}dB_t \\ dy & = & -(x - a + by)dt \end{array} \right.$$

The associated Fokker Planck equation is:

$$\frac{\partial P}{\partial t} = -\frac{\partial}{\partial x}[(x - \frac{x^3}{3} + y)P] - \frac{\partial}{\partial x}[(x - a + by)P] + \frac{\sigma}{2}\frac{\partial^2 P}{\partial x^2}$$

In fig 2.11, the dynamics of x is shown for three different noise variance but with the same realization of the random process integrated with a stochastic Heun scheme. Although different noise variances are used, the irregularities are similar within the 3 traces.

The power spetrum density is then computed for different values of σ and the frequency F_0 at which this power is maximum is reported on fig 2.12 and 2.13. When the random process is initialized by different seeds, F_0 increases smoothly with σ whereas it increases by jumping from one plateau to another when the same noise realization is used.

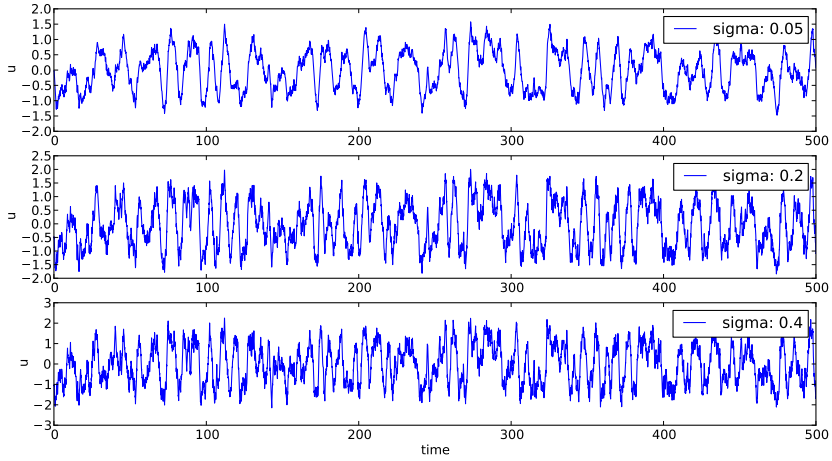


Figure 2.11: Examples of traces for various sigma.

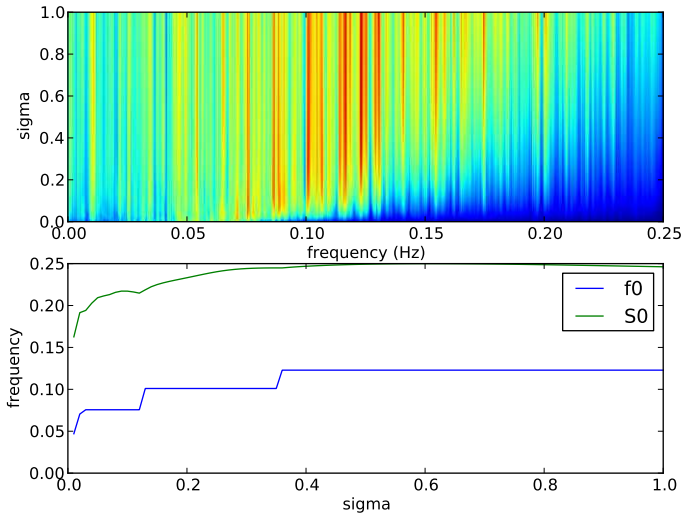


Figure 2.12: (Up) Spectrum for big values of the noise variance. (Down) Dominant frequency F_0 of the spectrum and energy S0 at this frequency. Noise realization are the same for all tested σ.

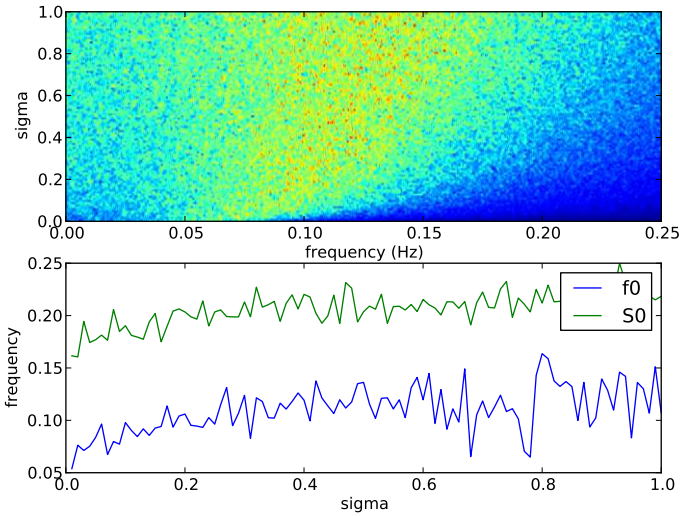


Figure 2.13: (Up) Spectrum for big values of the noise variance. (Down) Dominant frequency F_0 of the spectrum and energy S0 at this frequency. Noise realisation are different for all tested σ.

Lyapunov exponents calculated for different values of σ do not show chaotic dynamics as it stays negative on fig 2.14. In a wavelet analysis, bumps of high power are visible in two frequency bands. The low frequency component could result from noise induced switching between two stable fixed points inside the fold region and the high frequency component could be related to the limit cycle inside the Hopf region.

2.7 Hybrid dynamical systems.

In the Fitz Hugh Nagumo approximation of the membrane potential, the 2 dimensions dynamical system was smooth but the dynamics can also be reduced with a combination of smooth and discontinuous dynamics, with an instantaneous reset from the spike threshold to the resting potential. Such systems combining continuous and discontinuous dynamics are referred as hybrib systems [95].

2.7.1 Integrate and fire neuron models.

Leaky integrate and fire neuron (LIF): constant input. The simplest neuron model after the simple Poisson process consist of the membrane equation with a leak current and external input, spikes are generated by a discontinuity so that the membrane potential is reset to V_r when the voltage crosses a threshold

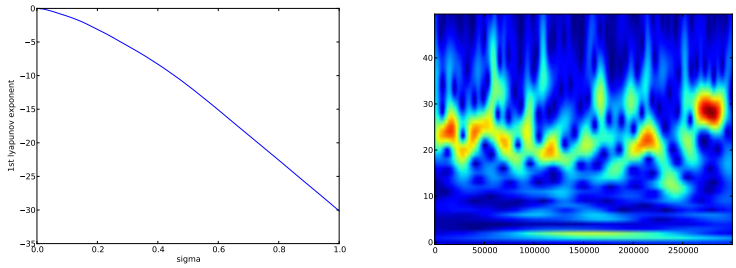


Figure 2.14: (Left) Lyapunov exponents depending on sigma. (Right) Time-frequency analysis of the signal with time steps on the horizontal axis and frequency in Hertz on the vertical axis.

value V_{th}. The dynamics then follows:

$$C\frac{dV}{dt} = -g_L(V - E_L) + I_{ext}$$

$$\text{If } V(t^-) > V_T \text{ then } V(t^+) = V_r.$$

In the case of constant external input I_0 applied by an electrode to an isolated neuron with no synaptic inputs, there is a threshold current $I_T = g_L(V_T - E_L)$ so that if $I < I_T$, the stable fixed point solution is the subthreshold potential $V_\infty = E_L + \frac{I_0}{g_L}$ and if $I > I_{th}$, the neuron spikes regularly. The time dependent solution of the equation is $V(t) = V_\infty + Ae^{-\frac{t}{\tau}}$ with $A = V(0) - V_\infty$ and the time constant, $\tau = \frac{C}{g_L}$. Considering the potential initially at its reset value, $V(0) = V_r$, the threshold is reached at time T so that $V_{th} = V_\infty + (V_r - V_\infty)e^{-\frac{T}{\tau}}$ and thus the stationary interspikes interval of the neuron is

$$T = \tau ln\frac{V_T - V_\infty}{V_r - V_\infty}$$

.

2.7.2 Diffusion approximation of Poissonian input.

In the Stein model [96], the free membrane potential (without considering the threshold for spikes) follows the stochastic differential equation (considering the normalized voltage $V \to g_L(V - E_L)$:

$$dV(t) = -V(t)dt + s_E dN_E(t) + s_I dN_I(t)$$

where N_E,N_I are Poisson processes of rates λ_E, λ_I simulating incoming spike trains and $s_E > 0$,$s_I < 0$ the amplitude of excitatory and inhibitory synaptic events. The diffusion approximation consist in taking simultaneously the limits of small amplitude of synaptic events and large rates of the Poisson processes,

it was shown in [97] that this model converges in law to the following Orstein-Uhlenbeck process:

$$dV(t) = (-V(t) + \mu)dt + \sigma dB(t)$$

with

$$\mu(t, v_0) = v_0 e^{-t} + (s_E \lambda_E + s_I \lambda_I)(1 - e^{-t})$$

and

$$\sigma^2(t, v_0) = \frac{s_E^2 \lambda_E + s_I^2 \lambda_I}{2}(1 - e^{-2t})$$

the drift and diffusion coefficients. The stationary density for the membrane potential is the following Gaussian probability distribution:

$$p(V, t|v_0, 0) = \frac{1}{\sqrt{2\pi\sigma^2(t, v_0)}} e^{-\frac{(V-\mu(t,v_0)^2}{2\sigma^2(t,v_0)}}.$$

2.7.3 Fokker-Planck equation.

We consider the LIF model:

$$C\frac{dV}{dt} = -g_L(V - E_L) + I_{ext}$$

$$\text{If } V(t^-) > V_{th} \text{ then } V(t^+) = V_r$$

with a refractory period τ_r during which V is clamped to the reset value. The Fokker-Planck equation for the probability distribution of the membrane potential is:

$$\frac{\partial P(V,t)}{\partial t} = \frac{\partial (g_L(V - E_L) - \mu)P(v,t)}{\partial v} + \sigma^2 \frac{\partial^2 P(V,t)}{\partial V^2} + r(t)\delta(V - V_r)$$

with μ and σ taken from the diffusion approximation and $\delta(x) = \begin{cases} 1 & \text{if } x = 0 \\ 0 & \text{else} \end{cases}$.

It can be written in the form of a conservation law:

$$\frac{\partial P(V,t)}{\partial t} = -\frac{\partial J(V,t)}{\partial V}$$

with the probability current

$$J(v,t) = (-g_L(V - E_L) + \mu)P(v,t) + \sigma^2 \frac{\partial P(V,t)}{\partial V} + r(t)H(V - V_r)$$

with H the Heaviside function. Boundary conditions on the lower part are $lim_{V\to-\infty} P(V,t) = 0$ and $lim_{V\to-\infty} VP(V,t) = 0$ so that $\int_{-\infty}^{V_T} P(V,t)dV$ is finite. At the threshold, the condition is absorbing $P(V_T, t) = 0$ and the probability current through threshold is the firing rate $r(t)$ so that

$$\frac{\partial P(V_T,t)}{\partial V} = -\frac{r(t)}{\sigma^2}.$$

Taking neurons in the refractory state into account, the normalization condition writes: $\int_{-\infty}^{V_{th}} P(V,t)dV + \int_{t-\tau_r}^{t} r(s)ds$. The stationary firing rate for this model is shown in [98] to check:

$$r^{-1} = \sqrt{\pi} \int_{\frac{V_r-\mu}{\sigma}}^{\frac{V_r-\mu}{\sigma}} e^{s^2}(1+\mathrm{erf}(s))ds$$

with erf the error function

$$\mathrm{erf}(x) = \frac{1}{\pi}\int_{-x}^{x} e^{-s^2}ds.$$

2.7.4 Non-linear integrate and fire model.

The leaky integrate and fire neuron model is simple to use but has limited behavior and several more recent models have integrated a non-linearity to simulate the spiking mechanism and a secondary variable to provide an adaptation mechanism. This non-linearity is quadratic in the Izhikevich model [99] and exponential in the Brette-Gerstner model [100] and with such sophistications, the diversity observed in neuronal dynamics can be reproduced easily. The Brette-Gerstner model is driven by the following equation:

$$C\frac{dV}{dt} = -g_L(V-E_L) + g_L\Delta_T e^{\frac{V-V_T}{\Delta_T}} - w + I$$

$$\tau_w \frac{dw}{dt} = a(V-E_L) - w$$

with C, g_L and E_L the same parameters as in the LIF, Δ_T shaping the spike, I the external input, τ_m, the adaptation time scale and a scaling the contribution of the voltage to the adaptive variable dynamics. When $V > V_T$, the exponential term grows very fast corresponding to a spike and when the voltage crosses V_{cut}, the voltage is reset and the adaptation variable is increased:

$$\text{if } V > V_{cut} \left\{ \begin{array}{l} V = V_r \\ w = w + b \end{array} \right.$$

When $I < I_{th}$, the system set in a quiescent fixed point and $I > I_{th}$ leads to persistent firing of the neuron through saddle-node bifurcation if $a < \frac{g_L}{\tau_w}$ and through Andronov-Hopf bifurcation if $a > \frac{g_L}{\tau_w}$. The bifurcation to persistent firing when $a = \frac{g_L}{\tau_w}$ corresponds to a codimension two Bogdanov-Takens bifurcation [101].

2.7.5 Parameters for excitatory and inhibitory neurons.

Inhibitory neurons are usually considered to be fast spiking cells, with no adaptation, and excitatory neurons shows adaptive behavior with their firing rate slowly decreasing when a constant input is injected. Parameters used in the

Parameters of the membrane potential		
Membrane time constant [ms]: $\tau_m = \frac{C}{g_L}$		20
Membrane capacity [nF]: C		0.2
Membrane resting potential [mV]: E_L		-60
Reset membrane potential [mV]: V_r		-60
Threshold membrane potential [mV]: V_T		-50
Cutting membrane potential [mV]: V_{cut}		20
Refractory period [ms]: τ_r		5
Parameters of the synapse		
Excitatory synapse time constant [ms]: τ_{synE}		2
Inhibitory synapse time constant [ms]: τ_{synI}		1
Excitatory synapse reversal potential [mV]: E_{syn_E}		0
Inhibitory synapse reversal potential [mV]: E_{syn_I}		-80
Parameters of the adaptation variable	**Regular spiking (E)**	**Fast spiking (I)**
Increment of the adaptation variable: b	0.04	0
Scaling of the membrane potential contribution to the adaptation dynamics: a	0.001	.1
Adaptation time constant [ms]: τ_w	120	50

Figure 2.15: Parameter used for excitatory and inhibitory cells in the adaptive exponential integrate and fire neuron model.

simulations are listed in fig 2.15. The parameters for the regular spiking cell are taken from the result of fitting procedure on Hodgkin-Huxley model [102] or real data [103]. The average firing rate of a neuron connected to a Poisson input spike train is plotted for various input frequencies in fig 2.16. The adaptation added to the excitatory neuron linearizes this frequency-response curve.

2.8 Columns.

As we saw in the introduction, columnar structure support the modular view of the brain. There are still some controversy about what is the definition of a column and its internal structure.

2.8.1 Definition.

Anatomical column. During cell migration, minicolumnar structures can be seen and will stay in such a packed form in the adult neocortex. These microcolumns have around $50\mu m$ diameter and contain from 80 to 100 neurons. It has been supposed from the 80's that it is a uniform structure across areas and species but more recent observations found inter-individual and inter-species variability in the size and density in neurons of these columns [78].

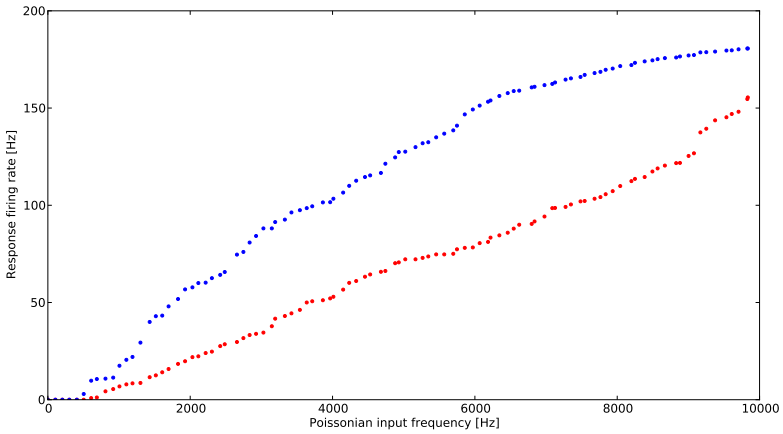


Figure 2.16: **Frequency-response curves.** - Average firing rate response to a Poisson input spike train at various frequencies for excitatory (red) and inhibitory (blue).

Functional column. At larger scale, as described in the introduction, functional columns are characterized by the response properties of its neurons. The diameter of a column is around $300\mu m$ and it is composed of around 60-80 minicolumns. Hypercolumns in primary visual cortex gathers cells having similar receptive field position but for which the preferred orientation may differ, it is around $1mm$ width. Columns related to more complex features can be found in inferotemporal cortex [104] or columns coding for a feature hold in memory can be found in prefrontal cortex [10].

2.8.2 Internal structure.

E-I network. The most common way to model a column is to consider a group of excitatory cells and a group of inhibitory cells. It is common to take 80% of excitatory cells and 20% of inhibitory cells. Composition of the network should also take the cell properties into account, with excitatory cells showing adaptation whereas inhibitory cells have fast spiking dynamics.

Neural circuit. A cortical column spans over 6 layers and networks accounting for this laminar structure are called neural circuits. In the Jansen and Rit model [105], three populations are considered for a column: one of excitatory pyramidal cells (located in layer II/III or layer V), one of inhibitory interneurons and one of excitatory interneurons located in IV. The LAMINART architecture also include three layers: II/III, IV and VI [106]. Detailed realizations of a neocortical column of the rat have also been realized including the detailed anatomy of the neuron in the blue brain project [80]. Templates based on the anatomical

studies of V_1 assessed the precise connection probabilities among the different layers and may offer a good description of the column (see [27] using data from [107] and [108]). More recent data describing a circuits of the visual cortex are available in [109].

2.9 Mean field equations.

Derivation of the Wilson-Cowan equations For a single column model, the activity is described by a macroscopic variable $x(t)$ describing the proportion of neurons firing at time t. After a neuron spikes, there is a refractory period τ_r during which the neuron is non responsive and for excitatory neurons, the proportion of neurons which are not in their refractory period at time t is

$$1 - \int_{t-\tau_r}^{t} X(t')dt'.$$

Neurons heterogeneity, introduced by a distribution $D(\theta)$ of firing threshold or a distrbution $C(w)$ on the number of afferent synapses to a neuron, shapes the response function, $S(x) = \int_0^x D(\theta)d\theta$ or $S(x) = \int_{\frac{\theta}{x}}^{\infty} C(w)dw$. This response can be taken as the sigmoid function, $S(x) = \frac{1}{1+e^{-a(x-\theta)}}$ whith the gain a and the threshold θ.

For a column with a level of recurrence α receiving the input β, the average excitation:

$$\int_{-\infty}^{t} h(t-t')(\alpha X(t') + \beta)dt'$$

with h an exponentially decreasing function. The dynamics then follow:

$$X(t+\tau) = (1 - \int_{t-\tau}^{t} X(t')dt')S(\int_{-\infty}^{t} h(t-t')(\alpha X(t') + \beta)dt'$$

By considering averages :$\bar{f}(t) = \frac{1}{s}\int_{t-s}^{t} f(t')dt'$, integrals can be approximated so that:

$$\int_{t-\tau}^{t} X(t')dt' \to r\bar{X}(t)$$

$$\int_{-\infty}^{t} h(t-t')X(t')dt' \to k\bar{X}(t)$$

By keeping X instead of $\bar{X}$ and using the Taylor formula $X(t+\tau) = X(t) + \tau\frac{dX}{dt} + o(\tau)$, we reach the following equation:

$$\tau\frac{dX}{dt} = -X + (1 - rX)S(\alpha X + \beta).$$

Bifurcations in a one excitatory population model Thus for a one population model with self-connection α and external input β, considering $r = 0$, the dynamics follows:

$$\frac{dx}{dt} = -x + S(\alpha x + \beta)$$

for which fixed points checks: $x_0 = S(\alpha x_0 + \beta)$. As shown on fig 2.17, there can be one or three fixed points depending on the parameters α, β. As the input β is varied, for sufficiently high values of the recurrence α, two saddle-nodde bifurcations occur when the linear part cancel out: $\alpha S'(\alpha x_0 + \beta) = 1$ which corresponds to the first diagonal $y = \frac{x}{\alpha}$ being tangent to the response function. As α is decreased, there is an α_c where the two saddle-node curves collide, this is the cusp of the system.

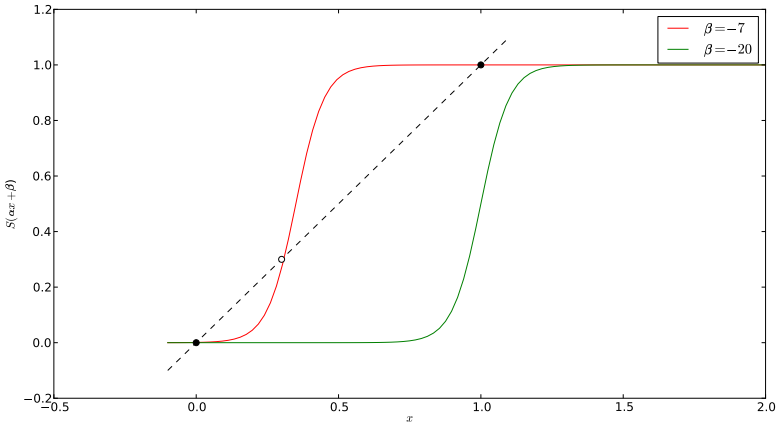


Figure 2.17: **Response function for various** β - There can be one or three fixed points depending on the number of intersection points between the response function and the line $g(x) = x$).

Phenomenological model of UP and DOWN states. Based on the previous demonstration of cusp bifurcation, the system can be approximated by its normal form near the bifurcation point. Near a cusp point, the population is described by the following dynamics: $\dot{x} = -\nabla E(r)$ deriving from the potential $E(r) = r^4 + ar^2 + bx$. When a goes from negative to positive values, the simple well potential becomes a double well potential and a scales the separation between the two fixed points and the height of the unstable fixed point. When the parameter $b = 0$, the double well potential is symmetric and one of the two fixed points has minimal potential when $b \neq 0$. The equation for the mean activity in a column is then:

$$dr = -(r^3 + ar + b)dt + \sigma dW_t.$$

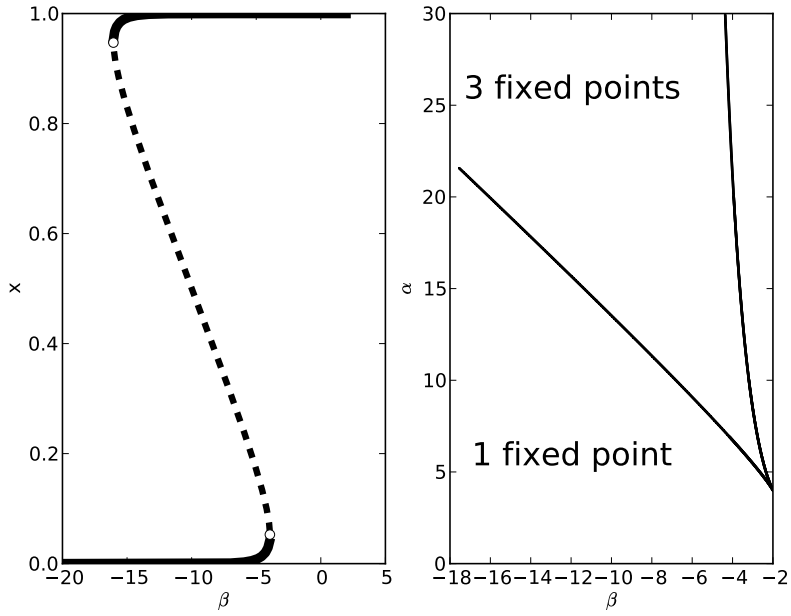


Figure 2.18: **Saddle-node bifurcation.** - (Left) SN bifurcations as β varies. (Right) Cusp point in the α, β.

When $a > 0$, the potential from which the dynamics derives has a single fixed point, $r_u = 0$, and the stochastic system has fluctuations around that fixed point whereas when $a < 0$, the corresponding potential has two fixed points, $r_s = \pm\sqrt{a}$, and transitions between these two fixed points are induced by noise. The corresponding Fokker-Planck equation is :

$$\frac{\partial p(r,t)}{\partial t} = \frac{\partial}{\partial r}(\nabla E(r) p(r,t)) + \frac{\sigma}{2}\frac{\partial^2 p(r,t)}{\partial r^2}$$

and the stationary distribution is:

$$p_0(r) = \mathcal{N} e^{\frac{-2\nabla E(r)}{\sigma}}.$$

The time between transition from one of the stable node to the other is given by the inverse of the Kramers rate [87]:

$$T_s = \frac{\pi}{|\sqrt{E''(r_u)}|E''(r_s)} e^{\frac{E(r_u)-E(r_s)}{\sigma}}.$$

If an additional low frequency, ω, forcing is added to the system, there is a noise intensity for which coherent transitions occurs between the two fixed points. The condition for such stochastic resonance is $2T_s = T_\omega$.

The description of the network activity in terms of attractors was used in a more sophisticated network modeling decision making in [110].

Bifurcations in the two populations model of a column. According to Dale's principle, neuronal cells should be considered based on their synaptic influence on other cells, into excitatory and inhibitory cells leading to the mean-field model:

$$\tau_e \frac{dE}{dt} = -E + (1 - r_e E) S(a_e(c_1 E - c_2 I - \theta_e + P))$$

$$\tau_i \frac{dI}{dt} = -I + (1 - r_i I) S(a_i(c_3 E - c_4 I - \theta_i + Q))$$

The system then have one or several fixed points or even limit cycles as shown in these classical results from Wilson and Cowan in [111]:

- If $c_1 > \frac{9}{a_e}$, there are some constant (P,Q) configurations for which the system has 3 fixed points.
- If $\frac{a_e c_2}{a_e c_1 - 9} > \frac{a_i c_4 + 9}{a_i c_3}$, there are some constant (P,Q) configurations for which the system has 5 fixed points.
- If $c_1 a_e > c_4 a_i + 18$, at least one fixed point is unstable.
- If $\frac{a_e c_2}{a_e c_1 - 9} > a_i c_4 + 9 a_i c_3$, $\frac{a_e c_1 - 9}{a_e c_2} < 1$ and the preceding condition for fixed point instability holds, then for $Q = 0$, there exists a threshold P_0 such by increasing P, a limit cycle appears when $P > P_0$.

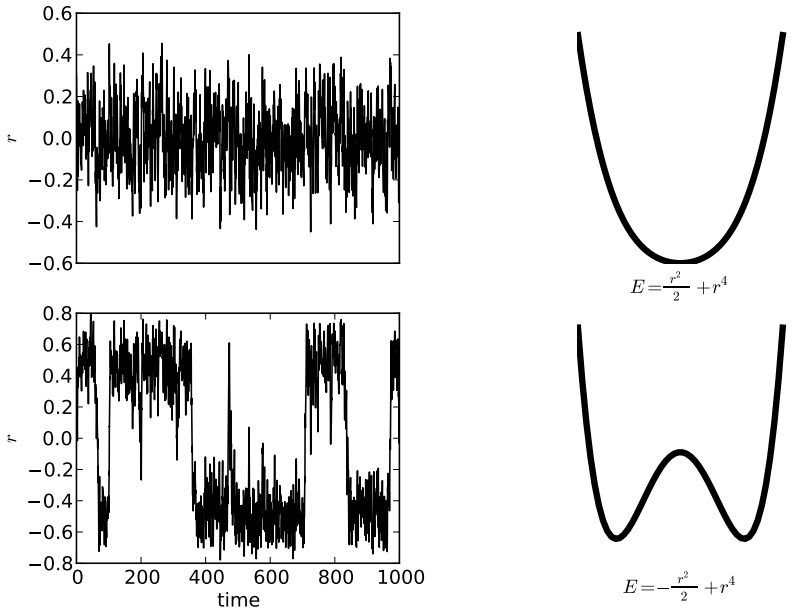


Figure 2.19: **Stochastic forcing of the normal form near the cusp bifurcation.** - (Up) $a > 0$, the firing rate r shows fluctuations around the single fixed point. (Down) $a < 0$, r has stochastic transitions between the two fixed points.

2.10 A column of spiking neurons.

The network is composed of two populations, 80% of the cells are excitatory and 20% are inhibitory. Excitatory cells are regular spiking with adaptation and inhibitory ones are fast spiking. When connections in the network are considered as sparse, there can be several behaviors depending on the balance between excitation and inhibition. For high values of the maximal excitatory conductance with low value of the inhibitory conductance, the network saturates to its maximal frequency and for higher values of the maximal inhibitory conductance, the network have collective oscillations or asynchronous state. The asynchronous irregular regime was first described in a theoretical work of Van Vreeswijk and Sompolinsky in [33] for a network of sparsely connected binary neurons with excitatory currents balancing precisely the inhibitory ones and this asynchronous state was later reported for integrate and fire neurons with current synapses [112] or conductance synapses [113].

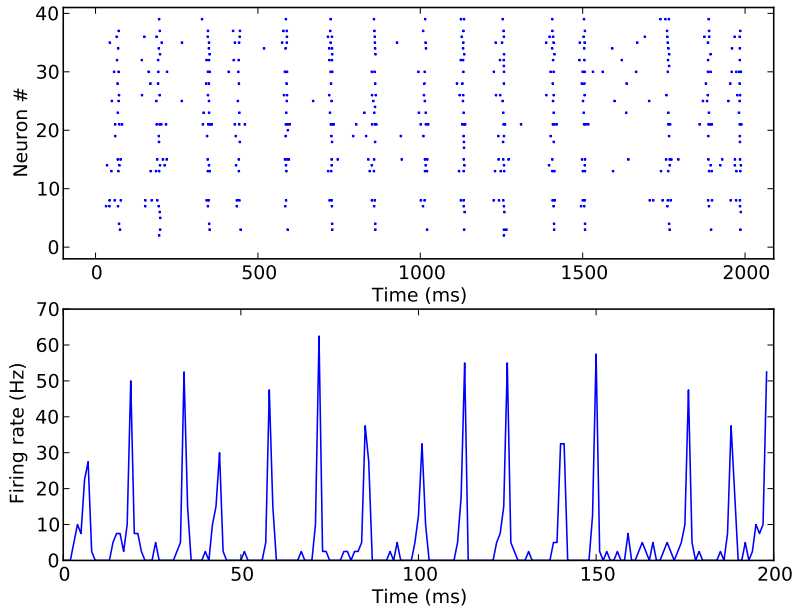


Figure 2.20: **Activity of the 40 excitatory neurons in a column composed of 50 neurons.** - (Up) Raster plot of a column. (Down) Average firing rate in the column.

Asynchronous irregular regime in a network of binary neurons. The network is composed of two populations, one with N_E excitatory neurons and

one with N_I inhibitory neurons. At each time, a neuron i from population k gets its state updated according to: $\sigma_i^k(t) = H(u_i^k(t))$ where H is the Heaviside function $H(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases}$ and $u_i^k(t)$ is the input to the neuron i at time t:

$$u_i^k(t) = \sum_{l=E,I} \sum_{j=1}^{N_k} J_{ij}^{lk} \sigma_j^l(t) + u_0^k - \theta_k$$

with u_0^k the external input and θ_k the threshold for neurons of the population k. Each neuron receive input from K neurons with connection strength $J_{ij}^{kl} = \frac{J_{kl}}{\sqrt{K}}$ so that the network is sparsely connected: $1 << K << N$. Moreover, because the absolute scale of the input is of no relevance, it is possible to consider $J_{EE} = J_{EI} = 1$ and $J_E = -J_{IE}$, $J_I = -J_{II}$. The population averaged firing rate is $m_k = \frac{1}{N_k} \sum_{i=1}^{N_k} \sigma_i^k(t)$ and it checks the mean field equation [33]:

$$\tau_k \frac{dm_k(t)}{dt} = -m_k(t) + \text{erf}(-\frac{u_k}{\alpha_k})$$

with the mean input for neurons of the population k:

$$u_k(t) = \sqrt{K}(\sum_{l=E,I} J_{lk} m_l(t) + E_k m_0) - \theta_k$$

where m_0 is the mean rate of external sources connecting with strength $\frac{E}{\sqrt{K}}$ to the excitatory population and $\frac{I}{\sqrt{K}}$ to the inhibitory population, and the variance of the input to the population is

$$\alpha_k(t) = \sum_{l=1,2} (J_{lk})^2 m_l(t).$$

The complementary error function is:

$$erf(z) = \int_z^\infty e^{-\frac{x^2}{2}} \frac{dx}{2\pi}.$$

Apart from saturating fixed points resulting in $m_E, m_I = 0, 1$ or $m_E, m_I = 1, 0$, there can be a balanced fixed with finite inputs so that:

$$Em_0 + m_E - J_E m_I = \text{O}(\frac{1}{\sqrt{K}})$$

$$Em_0 + m_E - J_I m_I = \text{O}(\frac{1}{\sqrt{K}})$$

so that as $K \to \infty$:

$$m_E = \frac{J_I E - J_E I}{J_E - J_I} m_0$$

$$m_I = \frac{E - I}{J_E - J_I} m_0$$

and the balanced state can exist if these stationary firing rates are positive, that is if:

$$\frac{E}{I} > \frac{J_E}{J_I} > 1$$

or

$$\frac{E}{I} < \frac{J_E}{J_I} < 1.$$

Furthermore there can be unbalanced solution $m_E, m_I = 1, 1$ with inputs $u_k = \sqrt{K}$ of order $\sqrt{K}$ if

$$\frac{E}{I} < \frac{J_E}{J_I} < 1$$

or if $1 > J_E$ and $1 > J_I$ so that the conditions to obtain a balanced state are:

$$\frac{E}{I} > \frac{J_E}{J_I} > 1$$

and

$$J_E > 1.$$

The balanced state achieve perfect tracking of a time varying input with the effective time constant being much smaller than for the unbalanced state.

Fokker-Planck equations for a column of integrate-and-fire neurons. A network with a similar column architecture was studied with integrate-and-fire neurons as units of the network [112]. The dynamics for each neuron is:

$$\frac{dV_i}{dt} = -\frac{1}{\tau}V_i + \sum_j \sum_k J_{ij}\delta(t - t_k - D)$$

and if

$$V_i(t) > V_{th}, V_i(t + \tau_{ref}) = V_r$$

where k runs over spikes of the neuron j and j runs over the input neurons to the neuron i consisting of C_{ext} neurons from the external source, $C_E = \epsilon N_E$ neurons from the excitatory population and $C_I = \epsilon N_I$ neurons from the inhibitory population with $\epsilon << 1$ so that connections are sparse. Moreover, notations are simplified by taking $C_I = \gamma C_E$. The delay D stands for the propagation time along the axon and the dendritic tree. The PSP amplitude for external and recurrent synapses are taken to be equal, J, and for the inhibitory synapses, the PSP amplitude is gJ. For such a network, the Langevin equation is:

$$\tau\frac{dV_i}{dt} = -V_i + \mu(t) + \sigma\sqrt{\tau}\eta_i(t)$$

with the average and variance of the input composed of a recurrent part and an external part, both resulting from Poissonian spike trains :

$$\mu(t) = C_E J(1 - \gamma g)\nu(t - D)\tau + C_E J\nu_{ext}\tau$$

and

$$\sigma = J\sqrt{C_E(1+\gamma g)\nu(t-D)\tau + C_E\nu_{ext}}.$$

The corresponding Fokker Planck equation for the probability density of the membrane voltage is

$$\frac{\partial P(v,t)}{\partial t} = \frac{\partial}{\partial V}((V-\mu(t))P(V,t)) + \frac{\sigma^2}{2}\frac{\partial^2 P(V,t)}{\partial V^2}$$

which can be rewritten in terms of probability flux:

$$\frac{\partial P(V,t)}{\partial t} = -\frac{\partial S(V,t)}{\partial V}$$

with

$$S(V,t) = -(V-\mu(t))P(V,t) - \frac{\sigma^2}{2}\frac{\partial P(V,t)}{\partial V}.$$

Boundary conditions should be precised for this equation to have a unique solution.

- At the threshold voltage, the probability flux gives the firing rate and this firing rate stays finite so that $S(V_{th},t) = \nu(t)$ and $P(V,t) = 0$ for $V \geq V_{th}$. This results in $\frac{\partial P(V_{th},t)}{\partial V} = -\frac{2\nu(t)\tau}{\sigma^2(t)}$.
- At the reset potential, the probability distribution is continuous and the probability flux from the threshold potential is reinjected taking the refractory period into account so that the probability flux has the following discontinuity $S(V_r^+,t) - S(V_r^-,t) = \nu(t-\tau_{ref})$ or

$$\frac{\partial P(V_r^+,t)}{\partial V} - \frac{\partial P(V_r^-,t)}{\partial V} = -\frac{2\nu(t-\tau_{ref})\tau}{\sigma^2(t)}.$$

- The integral of the probability distribution should stay finite so that $P(V,t) \to 0$ and $VP(V,t) \to 0$ when $V \to -\infty$.
- Finally, as a probability distribution, it checks the following normalization condition:

$$\int_{-infty}^{V_{th}} P(V,t)dV + \int_{t-\tau_{ref}}^{t} \nu(u)du = 1.$$

The stationary distribution solution of the Fokker Planck equation with such conditions for the voltage is

$$P_0(V) = \frac{2\nu_0\tau}{\sigma_0} e^{-\frac{(V-\mu_0)^2}{\sigma_0^2}} \int_{\frac{V-\mu_0}{\sigma_0}}^{\frac{V_{th}-\mu_0}{\sigma_0}} H(u-V_r)e^{u^2}du$$

. With H the Heaviside function, $H(x) = 1$ if $x > 0$ and $H(x) = 0$ otherwise, and

$$\mu_0 = C_E J\tau(\nu_{ext} + \nu_0(1-g\gamma)$$

$$\sigma_0^2 = C_E J^2 \tau(\nu_{ext} + \nu_0(1 + g^2\gamma)).$$

The normalization condition gives the stationary firing rate ν_0:

$$\frac{1}{\nu_0} = \tau_{ref} + 2\tau \int_{\frac{V_r-\mu_0}{\sigma_0}}^{\frac{V_{th}-\mu_0}{\sigma_0}} du e^{u^2} \int_{-\infty}^{u} dv e^{-v^2}.$$

Linear stability analysis gives the Hopf bifurcation lines where transition to synchronous spiking occurrs. The computation of the coefficient of variation of interspikes intervals on numerical simulations of the network then determines whether the spiking is regular or irregular. The diagram for the network with uniform delays $D = 1.5ms$ is shown on fig 2.21 depending on the external frequency and the relative strength of excitatory and inhibitory currents. There are three lines separating asynchronous behaviour from synchronous instabilities:

- A vertical line at $g = 4$ corresponds to exact balancing of inhibitory and excitatory currents. For $g < 4$, the activity is synchronous regular at high frequency and for $g > 4$, neurons spike at low frequency asynchronously.
- For $g > 4$, at low external input frequency, a branch separates the asynchronous irregular state from a low frequency synchronous irregular regime.
- For $g > 4$, at high external input frequency, a branch separates the asynchronous irregular state from a high frequency synchronous irregular regime.

Moreover, in the triangular region near $g = 4$ and with external frequency close to threshold, the activity combines a slow oscillation and a high frequency spiking on top of it.

2.11 Coupled columns.

2.11.1 Reduction to oscillators.

For two weakly coupled columns, the mean field equations are:

$$\begin{cases} \frac{dE_k}{dt} &= -E_k + S(A_k^e + \eta a_e U_l) \\ \frac{dI_k}{dt} &= -I_k + S(A_k^i + \eta a_i V_l) \end{cases}$$

with populations $k, l = 1, 2$ $k \neq l$ and

$$U_l = a_1 E_1 - a_2 I_l$$

$$V_l = a_3 E_l - a_4 I_l$$

$$A_k^e = a_e(c_1 E_k - c_2 I_k - \theta^e + P_k)$$

$$A_k^i = a_e(c_3 E_k - c_4 I_k - \theta^i).$$

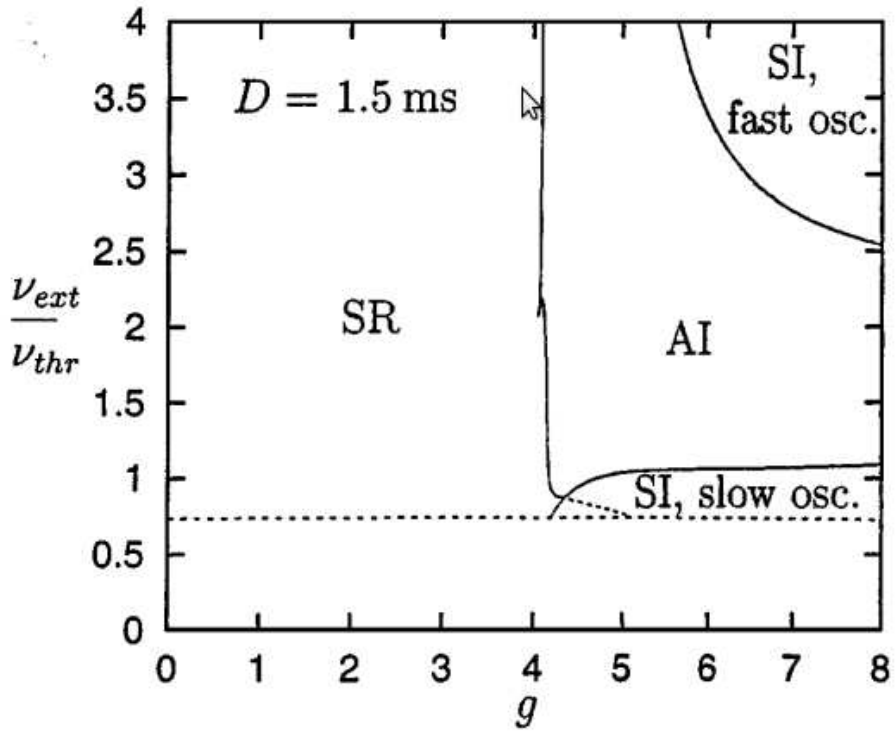


Figure 2.21: **Diagram of a network of IF neurons (Adapted from [112]).** - Parameters are the relative strength of inhibition g and the frequency of the external input rescaled by the frequency needed to reach threshold without feedback, $\nu_{th} = \frac{V_{th}}{C_E J \tau}$. Possible states are described as asynchronous (A), synchronous (S), regular (R) or irregular (I).

A change of variables leads to the Hopf normal form and the reduction of the oscillators to their phases ϕ_1, ϕ_2 leads to the system:

$$\begin{cases} \frac{d\phi_1}{dt} &= \omega_1 - K_{12} sin(\phi_1 - \phi_2) \\ \frac{d\phi_2}{dt} &= \omega_2 - K_{21} sin(\phi_2 - \phi_1) \end{cases}$$

with coupling terms proportional to $\eta a_e S'(A^e_k)$ and $\eta a_e S'(A^i_k)$.

2.11.2 Few coupled columns.

Arnold tongues. For a single oscillator with forcing frequency ω and self-coupling K:

$$\frac{d\phi}{dt} = \omega - K sin(\phi)$$

various mode locking are possible depending on ω and K. Arnold tongues are regions of the parameter space where the mode locking index $\frac{\dot{\phi}}{\omega}$ is uniform. The largest areas are integer modes and smaller areas are fractional modes, modes $0, \frac{1}{2}, 1, \frac{3}{2}, 2$ can be seen on fig 2.22.

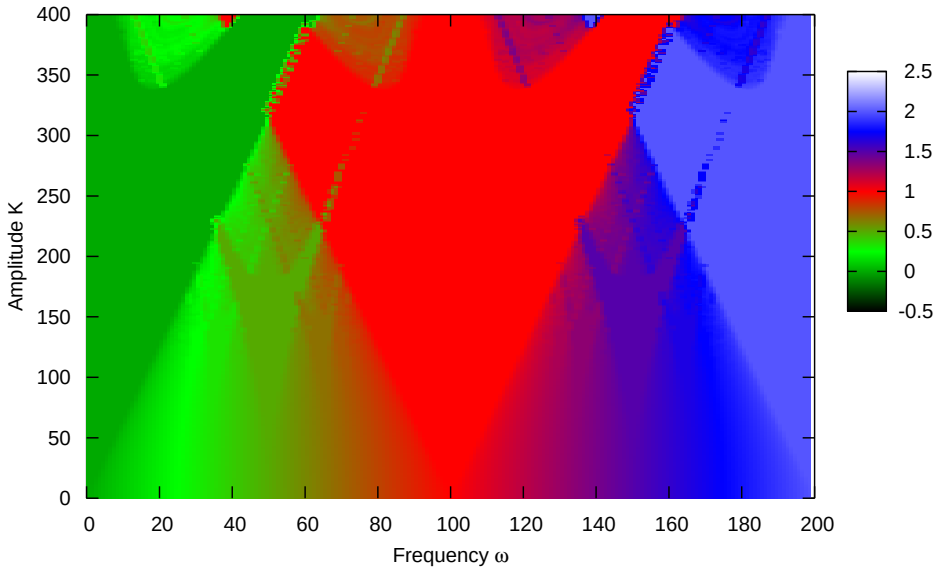


Figure 2.22: **Arnold tongues.** - Mode locking index $\frac{\dot{\phi}}{\omega}$ depending on the intrinsic frequency ω and the self-coupling K.

Synchronization transition. For two coupled oscillators forced at frequencies ω_1 and ω_2 and symmetrically coupled with strength K, it is convenient to consider the phase difference $\psi = \phi_1(t) - \phi_2(t)$ which follows:

$$\frac{d\psi(t)}{dt} = \Delta\omega + 2K\sin\psi$$

with $\Delta\omega = \omega_1 - \omega_2$. If the coupling is strong enough, the stationary phase difference is $\Delta\phi = \arcsin\frac{-\Delta\omega}{2K}$. If $K < K_c = \frac{\Delta\omega}{2}$, there is no stationary solution. There is thus a frequency synchronization transition at K_c and it is a second order transition as the order parameter is continuous and its derivative is discontinuous at the transition point. The transition can be seen on fig 2.23 where the average phase difference, $\Delta\phi = \frac{1}{T}\int_0^T \psi dt$, is plotted as a function of the coupling strength for oscillators frequencies, $\omega_1 = 1.2$ and $\omega_2 = 1$.

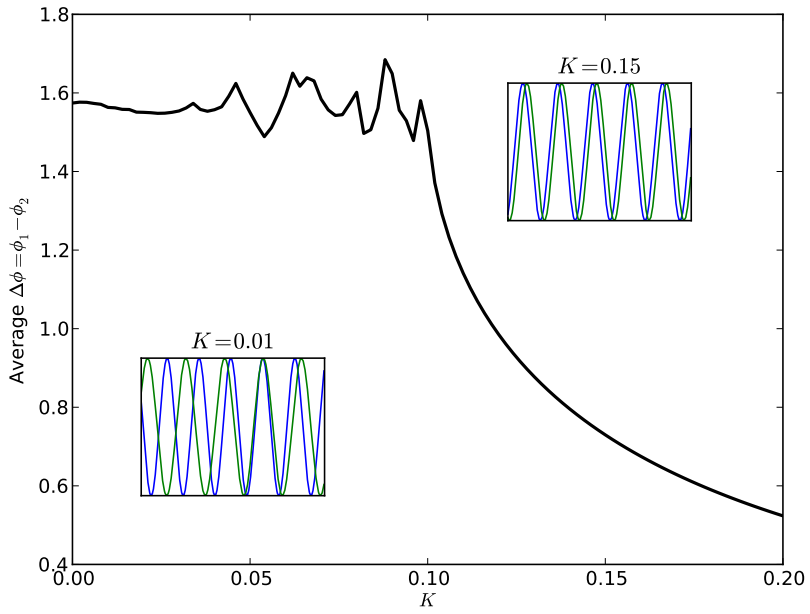


Figure 2.23: **Synchronization of two oscilators depending on the coupling K.** - The average phase difference is constant until K_c at which the two oscillators start getting synchronized.

Partial synchronization. With three oscillators, partial synchronization occurs as the coupling strength K is increased with oscillators with closest frequency synchronizing first. The sequence of such synchronization transitions is shown for three oscillators of frequencies $\omega_1 = 1$, $\omega_2 = 0.4$ and $\omega_3 = 0.2$ in fig

2.24 where the instantaneous frequencies of oscillators are plotted as a function of the coupling strength K.

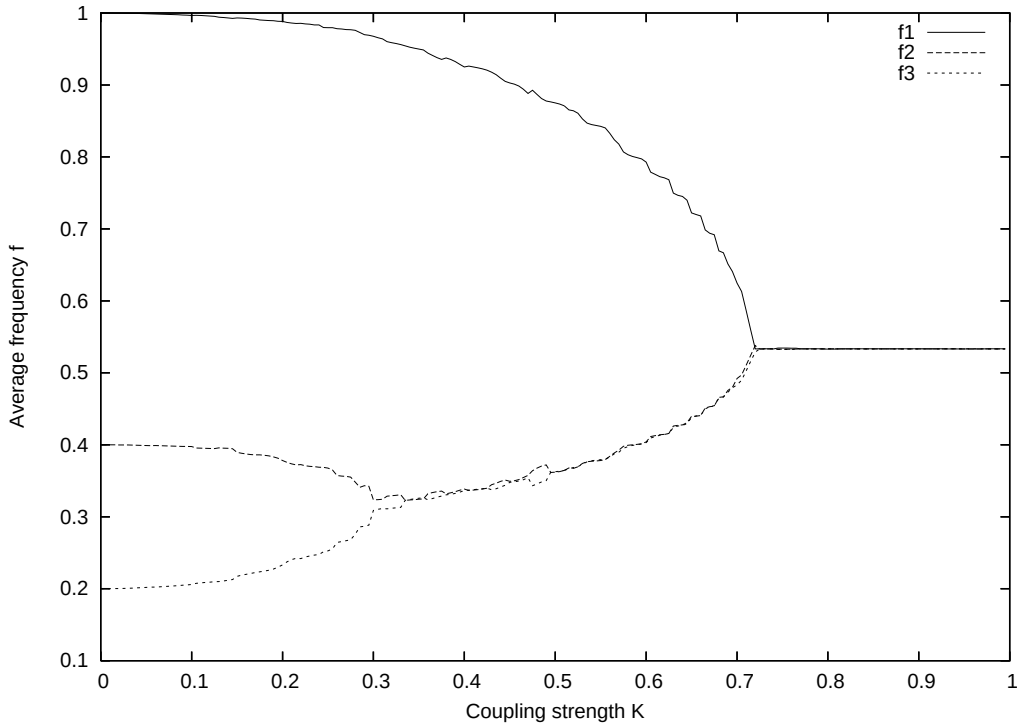


Figure 2.24: **Partial synchronization for 3 oscillators.** - For loose coupling, oscillators run at their intrinsic frequency, for intermediate values of the coupling only the two oscillators with the closest frequency are frequency locked and for higher values of the coupling, the network oscillates in full synchrony.

For $N \geq 4$, phase chaos was shown to occur through torus destruction as the coupling strength is increased [114].

2.11.3 Large population of coupled oscillators.

A population of globally coupled oscillators. The Kuramoto model is widely used to study synchronization between coupled units [115] and it is composed of N oscillators coupled, with $N >> 1$, according to the equations:

$$\frac{d\phi_i}{dt} = \omega_i + \sum_j K_{ij} sin(\phi_j - \phi_i)$$

with $1 \leq i \leq N$, intrinsic frequencies distributed according to $g(\omega)$ and the original model the coupling is homogeneous $K_{ij} = \frac{K}{N} > 0$. The dynamics can

then be written as:

$$\frac{d\phi_i}{dt} = \omega_i + Krsin(\psi - \phi_i)$$

with the order parameter defined as $re^{i\psi} = \frac{1}{N}\sum_k e^{i\phi_k}$, $r \in [0,1]$ measures the coherence in the population and ψ is the average phase. The coherence is $r = 0$ for weak coupling with oscillators moving independently and it is $r = 1$ when strong coupling makes all phases equal to ψ. Equivalently, the order parameter can be expressed a function of the probability distributions for intrinsic frequencies ,$g(\omega)$, and for the phases, $\rho(\phi, \omega, t)$:

$$re^{i\psi} = \int_{-\pi}^{\pi}\int_{-\infty}^{\infty} e^{i\phi}\rho(\phi,\omega,t)d\omega d\phi.$$

Moreover, the distribution ρ satisfies the continuity equation:

$$\frac{\partial\rho}{\partial t} + \frac{\partial}{\partial\phi}[(\omega + rsin(\psi - \phi)\rho]$$

and the normalization condition:

$$\int_{-\pi}^{\pi}\rho(\omega,\phi,t)d\phi.$$

The incoherent solution corresponds to $r = 0$ and $\rho = \frac{1}{2\pi}$ with all phases having equal probability of being occupied. A branch of partially synchronized solutions starts at $K_c = \frac{2}{\pi g(0)}$ and with a Loretzian distribution [9] for g, the coherence behaves as $r = \sqrt{1 - \frac{K_c}{K}}$.

Chimera states. We now consider a ring of oscillators with long range connections:

$$\frac{d\phi(x,t)}{dt} = \omega + \int_0^1 K(x,x')sin(\phi(x,t) - \phi(x',t) + \alpha)dx'$$

where the connections are made through a Gaussian kernel $K(x,x') = Ae^{-\frac{|x-x'|^2}{2\sigma^2}}$. With initial condition constant over an interval and randomly distributed with a Gaussian profile on its complementary, the network settle in a chimera state where a part of the network is phase locked and the other part is oscillating in an asynchronous manner as shown in fig 2.25.

Flip-flop network. A network of oscillators can be coupled to Wilson-Cowan units with equations:

$$\frac{dx_i}{dt} = -x_i + \sigma(cos(\phi) - cos(\phi_0)) + \sum_j w_{ij}f(x_j) + I$$

[9] A Lorentzian distribution is of the form $g(\omega) = \frac{\gamma/\pi}{\gamma^2+\omega^2}$.

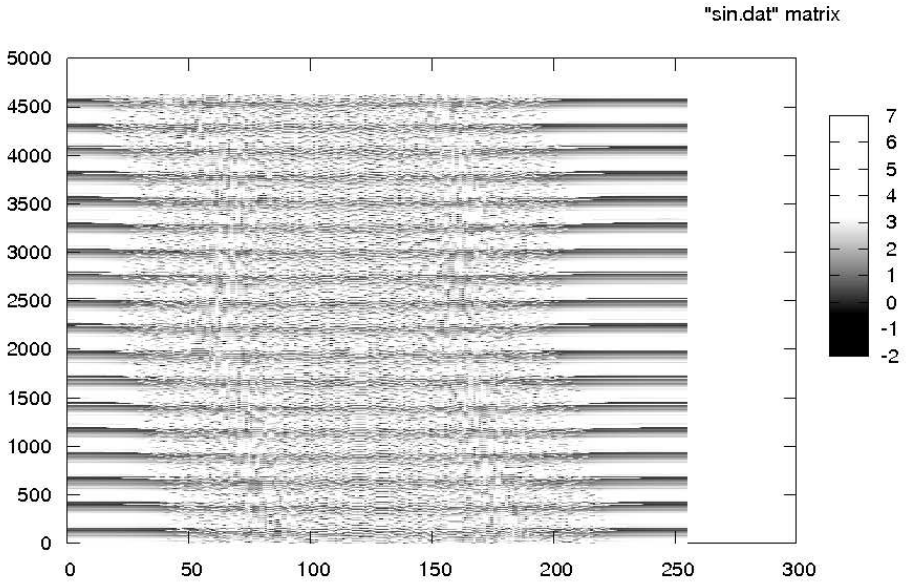


Figure 2.25: **Chimera state.** - Ring of oscillators with a part of the network phase locked and the complementary oscillating asynchronously.

$$\frac{d\phi_i}{dt} = \omega + (\beta - \rho x_i) sin(\phi_i)$$

where x_i is the membrane voltage of the component i and $cos(\phi_i)$ is an oscillating contribution to the dynamics scaled by σ and the membrane impacts the oscillators dynamics by a factor rho. $f(x_i)$ is the firing rate of the neuron i taken as an hyperbolic tangent and ϕ_0 is chosen so that the state where $(x_i, \phi_i) = (0, \phi_0)$ is a fixed point for the network. The conditions for stability of this fixed point are given in [116] reproduced in appendix. Two coupled units show down state, low frequency, antiphase oscillation for weak coupling and up state, high frequency, in-phase oscillation for large coupling. Before the phase-locked solution exists, a small window of chaotic behaviour is observed. In a larger network, the spontaneous activity wanders around cell assemblies storing memories.

2.12 Conclusion.

In this chapter, we explored the dynamics of computational units of the brain using bifurcation theory and the Fokker-Planck equation. The Hodgkin Huxley model for the dynamics of the membrane potential of a neuron is difficult to analyze and heavy to integrate. The FitzHugh Nagumo model is a reduction of the dynamics to a two dimensional phase model controlled by three parameters. With the analysis of bifurcations in this system, we found an organizing center for the dynamics in the neighborhood of which any possible dynamics is accessible. Stochastic forcing at this point resulted in the emergence of multiple timescales which may be traces of the attractors available in its neighborhood: limit cycle for the fast timescale and stochastic transitions between two fixed points for the slow timescale. We also noticed, for identical realizations of the noise, that the formant of the dynamics increases in a plateau like fashion when increasing the variance of the noise. The integrate and fire model is another phenomenological model commonly used for efficient simulation and we introduced a two dimensional version, the adaptive integrate and fire neuron which have a large repertoire of dynamics. For a column, mean field models provides a compact description and the network activity can be characteried by the synchrony and the regularity of firing. When columns have a collective oscillation resulting from a Hopf bifurcation in their mean field equation, the rich variety of possible dynamics was described using networks of coupled phase oscillators: resonance, frequency synchrony, phase synchrony and chimera states.

We saw that the asynchronous state described in the first part can be modeled as the chaotic dynamics of a balanced network of sparsely connected binary neurons. Networks of spiking neurons have a similar state and it can even be self-sustained for a very long time with conductance based neurons. The length of these supertransients of irregular activity depends exponentially on the size of the network [34]. The computation of Lyapunov exponents for hybrid dynamical systems needs special care [117] and when the largest of these exponent is not positive, this state is called stable chaos. Chaotic behavior of a macroscopic

variable, collective chaos, and chimera states have also been reported recently in networks of spiking neurons [118]. Neuronal networks thus have rich variety of dynamics and their potential use for solving computational tasks offers new approaches in artificial intelligence [119].

Chapter 3

Models of the formation of V1 and of its dynamics.

3.1 Models of the formation of V1

3.1.1 Introduction

Several functional properties of the brain are represented as maps: the body is represented by motor and somatosensory homunculus and the surrounding space is represented in cognitive maps of the hippocampus. The primary visual cortex is also endowed with multiple feature maps which provide a representation of the input space. Retinotopy is the mapping of spatial positions of the visual space onto the surface so that close points of the visual space are represented by neurons which are close together in the cortex. Ocular dominance of a neuron indicates how its inputs are biased toward the left or right eyes and this selectivity is organized into alternating bands. The selectivity of neurons to orientation also forms maps with special points, pinwheel singularities, around which the preferred orientation varies smoothly. Those maps are structurally related, with lines of iso-orientation orthogonal to the frontiers of ocular dominance domains. Some other features are also encoded in V1 like spatial frequency or direction of motion. In an early model proposed by Hubel and Wiesel, orientation selectivity result from the specific pattern of connections from LGN inputs but later studies highlighted the contribution of lateral connections for sharpening the orientation tuning curve.

Hebbian plasticity rule on the feedforward and lateral connections leads to the formation of orientation maps in firing rate models. Obtaining similar results with spiking neurons and spike timing dependent plasticity rules is still an open challenge and require heavy computational resources. On the other side, firing rate models are only a coarse grain version of the dynamics and don't have the fine scale complexity of spiking neurons. The column based networks of spiking we propose combine these two approaches for efficient simulation of the

visual cortex. A grid of columns of spiking neurons having the same structure as exposed in chapter 2 are used to study the fine temporal dynamics of the network after a transient stimulation and a firing rate model is used for the learning an orientation map, each unit representing a column and the connection weight between two units representing the probability of connection between two neurons of their corresponding neurons.

This chapter aims at describing the ongoing activity in models of V1, its influence on visual information processing and its structuration through learning.

Four approaches to the formation of V1 are presented and we present results on three large scale implementations of V1 models. First, in an ice cubes model of a pinwheel inspired, connections are hard wired to implement orientation selectivity and we study ongoing dynamics and the response to static or rotating bars. A second model is a large grid of columns with isotropic connection kernels for which the ongoing dynamics and the response to focalized stimulation are studied. The connections in the third network are the outcome of Hebbian learning when the connectivity of the isotropic model is used to initialize connection weights so that we can show how visual experience shapes spiking correlations in the network.

3.2 Ice cubes model.

3.2.1 Description of the model.

The simplest way to build a model of the visual cortex is to pack columns together in a "crystal-like" manner so that it achieves its function of local extraction of oriented lines. The construction of V1 in a hierarchical manner was initially proposed by Hubel and Wiesel in the 60's [82] based on their electrophysiological recordings in the cat area 17. The redundancy of coding that they discovered in the vertical direction (cortical depth) suggested that the cortical column could be the unitary building block of the cortex. The model of V1 they proposed is then obtained from an appropriate packing of these elementary units with hard wired connections. The wiring of these fixed connections is derived from the observed cortical fibers and from simplifying hypothesis when these observations are not possible.

As connections are fixed, retinotopy is imposed by simply providing a set of neighboring cortical columns with afferent connections from cells corresponding to a defined position. This group of columns defines an hypercolumn. The whole retinal space is then represented by an array of such hypercolumns, see fig 3.1. The ocular dominance domains can also be modeled by alternating an hypercolumn taking its inputs from the right eye with an hypercolumn taking its inputs from the left eye.

The orientation selectivity is implemented by considering a cylindrical parametrization (r, θ) of the hypercolumn. Columns having a similar azimuth θ code for the same orientation $\phi = \frac{\theta}{2}$ and the radial dimension is redundant. The topological charge of the pinwheel is positive if orientations are turning clockwise

and negative for anti-clockwise rotation. More detailed models consider that the r-dimension is used to code the selectivity to spatial frequencies [120]. The shape of the receptive field of simple cells could then be explained by a linear summation of aligned LGN inputs from ON and OFF channels, see fig 3.2. The resulting receptive field is a derivative of Gaussian and gives an effective description of how a cortical cell from V1 receives inputs from the eyes through the LGN. Other types of cells, like complex (phase invariant) or hypercomplex (tuned to an optimal bar length), can be built on the same principle with a pool of simple cells with same position, same orientation and all possible phases connecting to complex thus being phase invariant. This is an example where the serial and hierarchical doctrines of the with feedforward projections from the eye to the visual cortex is very efficient.

The lateral connectivity in primary visual cortex is classically modeled by a difference of Gaussians that is excitatory at short range and inhibitory at long range although, as described in the introduction of this chapter, more realistic architecture should include long range excitation.

3.2.2 Parameters of the model.

A pinwheel is composed of $N \times N$ columns placed on a regular grid with coordinates $X \in [-1, 1]$ and $Y \in [-1, 1]$. Each column receives afferent inputs from all the cells in the retinal layer and in this model all columns code for the same location, columns only differ in their preferred orientation. The retinal layer is also composed of 11×11 cells with coordinates $x \in [-1, 1]$ and $y \in [-1, 1]$. The activity of these retinal cells is modeled as inhomogeneous Poisson processes with rates corresponding to the light intensity of the visual input.

The afferent connectivity of a column c located at (X_c, Y_c) in the cortical plane with azimuth $\theta = \arctan \frac{X_c}{Y_c}$ from a cell located at a retinal position (x, y) is given by a directional second derivative of Gaussian. For $X_c = 0$, the second derivative of $G(x, y) = e^{-\frac{x^2+y^2}{\sigma^2}}$ is taken along the x-dimension:

$$\frac{\partial^2 G(x, y)}{\partial x^2} = A_0(1 - \frac{2}{\sigma^2}x^2)e^{-\frac{x^2+y^2}{\sigma^2}}$$

with A_0 a constant. For $\theta \neq 0$ and taking $A_0 = 1$, the connectivity is obtained by considering a rotation of the coordinates:

$$W^{aff}_{xy,\theta} = (1 - \frac{2}{\sigma^2}(x \cos\frac{\theta}{2} + y \sin\frac{\theta}{2})^2)e^{-\frac{(x\cos\frac{\theta}{2}+y\sin\frac{\theta}{2})^2+(-x\sin\frac{\theta}{2}+y\cos\frac{\theta}{2})^2}{\sigma^2}}$$

where σ parametrizes the sharpness of the receptive field. As density of connections, it is restricted to $[-1, 1]$, with negative weights corresponding to the inhibitory projections and positive weights to the excitatory ones. The integral of this function is zero so that a column should receive as many excitatory connections as inhibitory connections. In practice, as the retinal space is restricted to $[-1, 1] \times [-1, 1]$ and most inhibitory connections are too small to make any

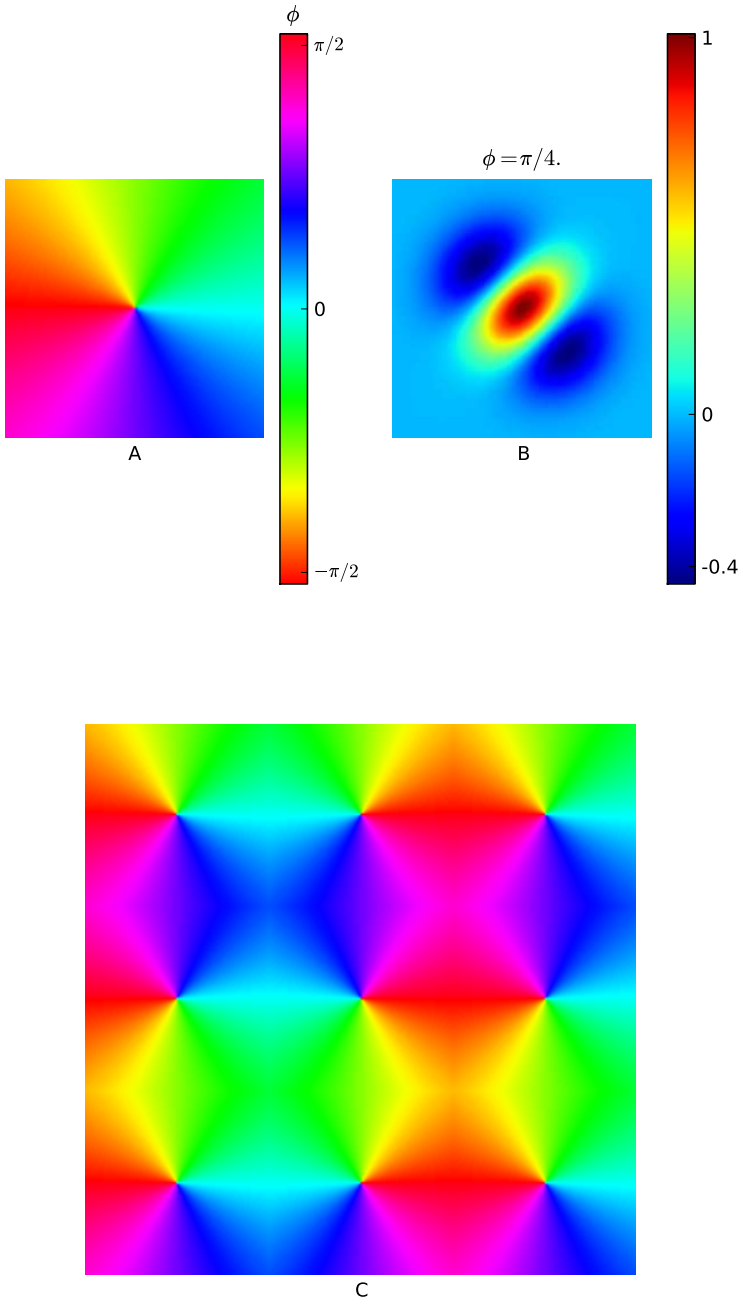


Figure 3.1: **Ice Cube model** - (A) A pinwheel with colors representing the preferred orientations of columns turning clockwise. (B) The receptive field of a column having $\frac{\pi}{4}$ as preferred orientation. (C) A "crystal-like" orientation map is built by packing pinwheels together

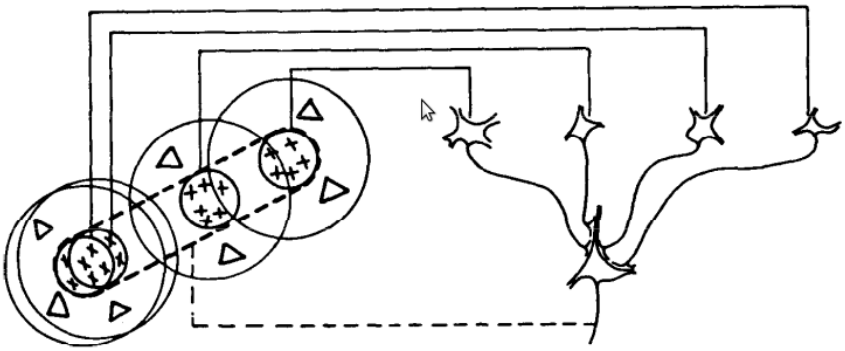


Figure 3.2: **Simple cell model.** - Cells from the LGN with ON center receptive fields feeds giving its orientation preference (Adapted from [82]).

connection [1] and the excitatory connections slightly overtake the inhibitory ones.

The lateral connectivity is restricted to the nearest neighbors, the column has a density $p = 0.2$ of self connections and the 8 neighboring column project to it with probability $p = 0.1$ both for excitation and inhibition. Parameters of the neurons are the same as described in table 2.15 of Chapter 2, a column being composed of 40 excitatory cells and 10 inhibitory cells.

During the 500 ms before the presentation of a stimulus, the cortical layer is ignited with a Poissonian stimulation at high rate, each column is connected with density 0.2 to a 5000 Hz Poisson spike train. The visual stimulation is a static or a rotating bar and the retinal activity corresponding to this stimulation is a set of Poisson spike trains with rates:

$$a_{xy} = f_0 e^{-(\frac{x\cos\alpha(t)+y\sin\alpha(t)}{\sigma_L})^2+(\frac{-x\sin\alpha(t)+ycos\alpha(t)}{\sigma_l})^2}$$

with $\alpha(t)$ the orientation of the bar,f_0 the maximum frequency $60Hz$, $\sigma_L = 0.35$ the length of the bar and $\sigma_l = 0.05$ the width of the bar.

3.2.3 Analysis of the dynamics.

On-going activity: Phase diagram The maximal conductances of excitatory (g_E) and inhibitory (g_I) is varied and the resulting mean firing rate, mean coefficient of variation of interspikes intervals (ISI) and mean pairwise correlation between spike trains are plotted in a phase diagram on fig 3.4. There are three distinct regimes:

- When $g_I = 0$, the network is highly active with regular spiking of the cells but spikes are asynchronous.

[1] For two columns A and B, each containing N neurons, for a connection probability $P_{AB} < \frac{1}{N^2}$, there will be nearly no connection from neurons of A to B.

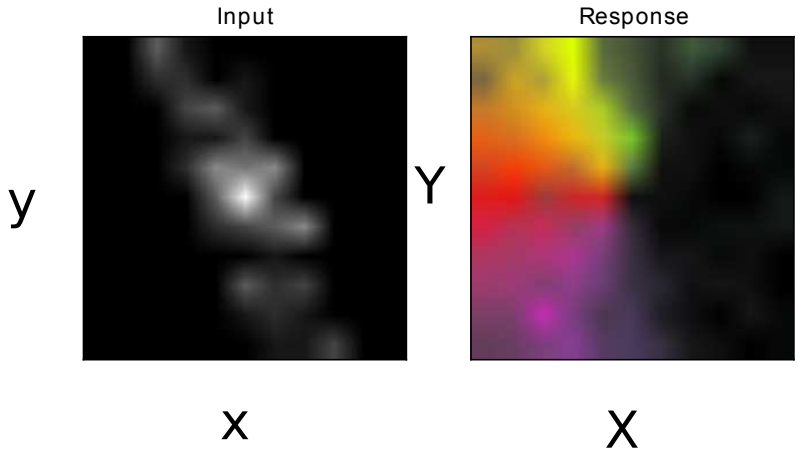


Figure 3.3: **Response of the pinwheel to an oriented bar** - (Left) Retinal activity corresponding to an oriented bar. (Right) Response of the pinwheel, the contrast indicates the firing rate of the column and the color codes for the preferred orientation of the column.

- When $g_E > g_I$, currents are dominated by excitation, the network is highly active with correlated regular spiking. This shows the synchronizing effect of inhibition.
- When $g_I > g_E$, currents are dominated by inhibition, the network activity is low and decreases when g_I is increased. Spiking occurs in an asynchronous irregular fashion.

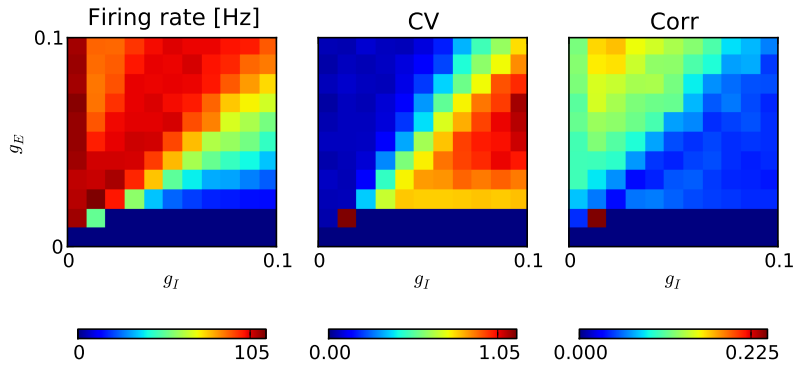


Figure 3.4: **Phase diagram for the Ice cubes model** - (Left) Average firing rate in Hertz. (Middle) Average coefficient of variation of the ISI distribution of spike trains. (Right) Average local correlation, that is correlation between spike trains of neurons belonging to the same column.

Response to a rotating bar. A rotating bar is presented to the network with a speed of 60Hz. A snapshot of the response of the network in the inhibition dominated regime is shown on fig 3.5. The activity in the inhibition dominated regime is a bump rotating at the same speed as the rotating bar, even at high rotation frequency [2], and this periodic activity is shown on fig 3.5. In the excitation dominated regime, no response specific to the stimulus can be seen.

[2] By the discretness of the grid, at very high frequency rotation, the bar is flashed periodically rather than rotating smoothly.

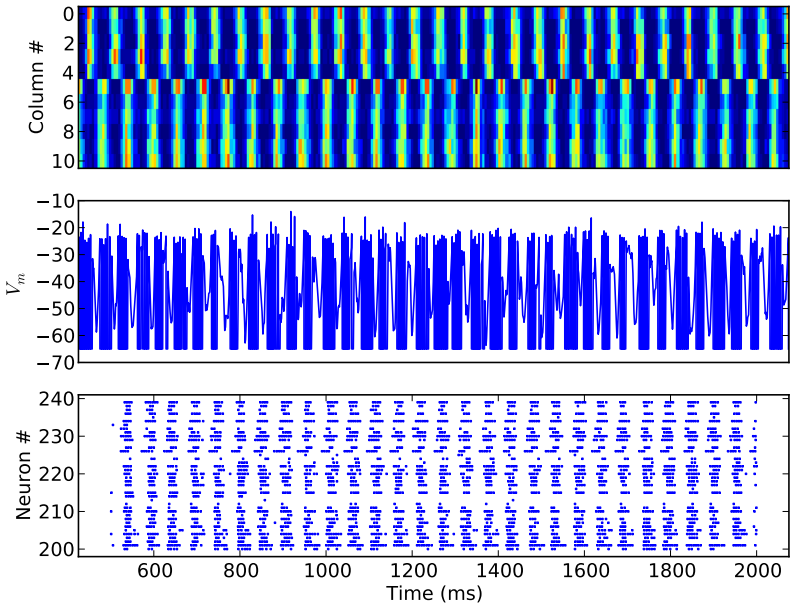


Figure 3.5: **Response to a rotating bar** - (Top) Average firing rate for a line of columns passing through the center. (Middle) Membrane potential of a neuron in column at position (0,5). (Bottom) Raster plot of the column at position (0,5).

Orientation tuning curve. As can be expected from the response to a rotating bar, the response to static bar is also very different depending on the regime of spontaneous activity. For characterizing the functional response of a cell, it is common to use the tuning curve that measures the firing rate of a cell as a function of the orientation of the presented stimulus ϕ. In our case we record the tuning curve of two columns at different positions of the pinwheel, one is close to the pinwheel singularity and the other is far (five columns away) from the pinwheel singularity. For each situation the map is shifted so that the central column of the network is the one from which the tuning curve is recorded. When the network is in the excitation dominated regime, the tuning curve is flat because the response is washed out by the spontaneous activity. In the inhibition dominated regime, the tuning curve is peaked with the maximum indicating the preferred orientation as can be seen on fig 3.6. In this regime, the tuning curve is much sharper for a column close to the pinwheel center than for a column far from the pinwheel center and the level of correlations is the same for all orientations, see fig 3.7. The level of activity in neighboring columns explains the sharpness of the orientation tuning curve close to the pinwheel center. Those columns around a column have very preferred orientation close to the pinwheel singularity and have low firing rate and thus do not excite the central column whereas far from the pinwheel, neighboring columns are excited for orientations similar to the preferred orientation of the central column.

3.3 Phenomenological models

The Ice Cubes model of $V1$ is based on the observations gathered from biological studies of the connectivity patterns, it tries to reproduce by hand these connectivity patterns and adds minimal assumptions when biology cannot help in choosing between different options. We explore here the models which forget about the biological processes but focuses on more abstract properties of $V1$. They aim to answer to simple questions like "What is the goal to achieve during the formation of $V1$?" or "What are the universal properties governing the formation of $V1$?" and can be analyzed in some simple cases. Such models are reviewed in [121].

3.3.1 The Elastic net theory

Multiple features are engrafted in V_1 through its functional maps for retinotopy, occular dominance and orientation preference. The mapping of visual stimulations parameters (X and Y positions, x and y coordinates of the preferred orientation and occularity) onto the cortical surface in layer IV is achieved during the formation of $V1$ with points close in the parameter space falling close onto the cortex. With the Elastic Net model, the formation of $V1$ is seen as an optimization process directed toward this mapping with the minimal wiring length.

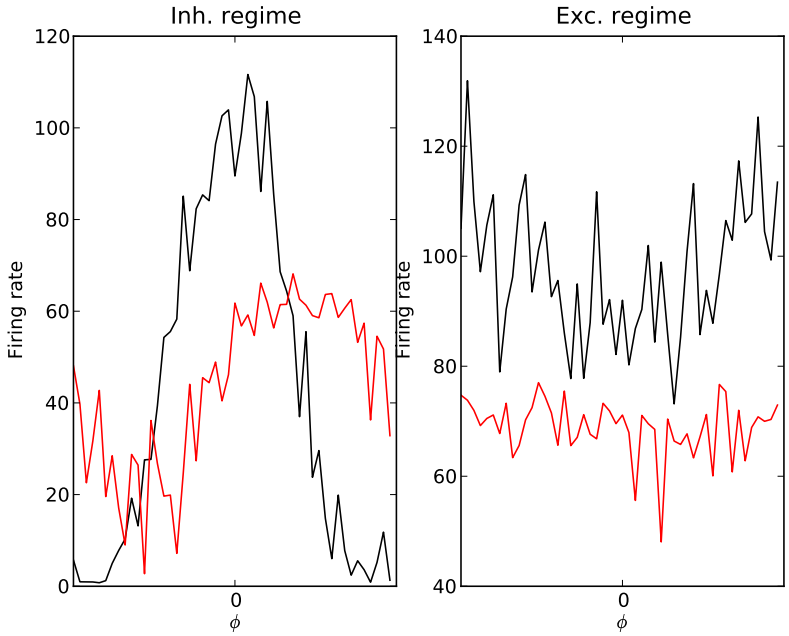


Figure 3.6: **Orientation tuning curves.** - (Left) Inhibition dominated regime, close to the pinwheel center in black and far in red. (Right) Excitation dominated regime with the same color conventions.

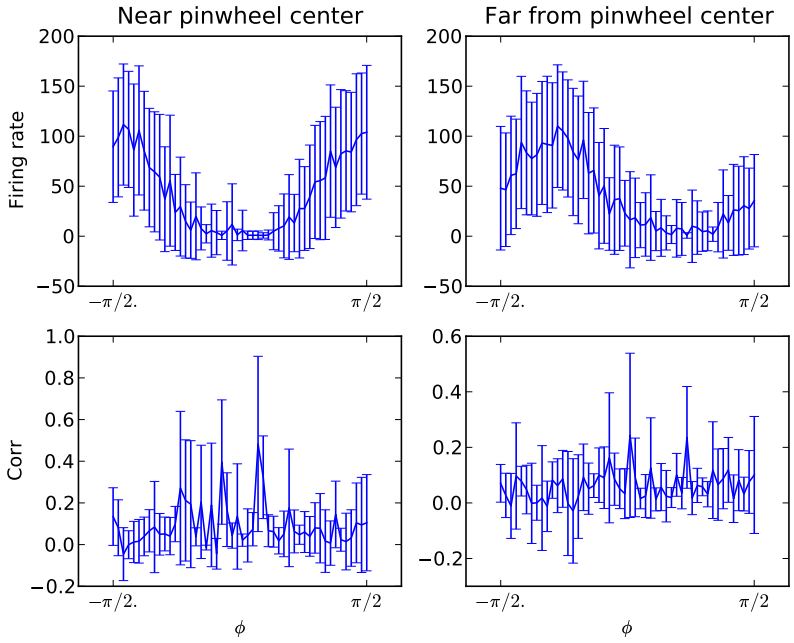


Figure 3.7: **Orientation tuning curves and correlations.** - (Left) Firing rate and correlations for a column close to the pinwheel center. (Right) Firing rate and correlations for a column far from the pinwheel center.

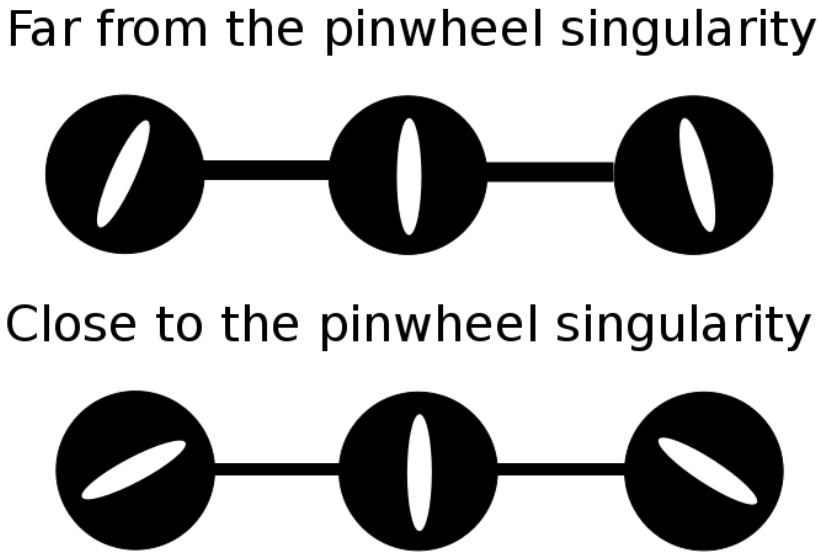


Figure 3.8: **Neighboring columns** - Preferred orientation and activity level (thickness of links) of neighboring columns.

A simple framework to apply the Elastic Net algorithm is the traveling salesman problem (TSP). A salesman has to visit N cities placed on a 2-dimensional plane and the problem is to find the shortest closed tour in which all cities are visited. This problem is NP-complete, that is it cannot be solved in a polynomial time of the number of cities to visit. An exhaustive search is possible when the number of cities is small but the computation time becomes very large as N grows. The TSP is formally equivalent to the mapping of N points regularly spaced on an elastic ring onto N points of the plane.

There are N cities with positions $(\mathbf{x_i})_{1\leq i\leq N}$ on the plane and N path points with positions $(\mathbf{y_j})_{1\leq j\leq N}$ on the elastic ring. The rule for changing the position of a path point is the following:

$$\Delta\mathbf{y_j} = \alpha\sum_i w_{ij}(\mathbf{x_i}-\mathbf{y_j}) + \beta k(\mathbf{y_{j+1}} - 2\mathbf{y_j} + \mathbf{y_{j-1}}),$$

where α scales the contribution of the cities in attracting path points and βk scales the elastic forces from the neighboring points on the path. The weight w_{ij} represents the normalized attraction that the city point i has on the path point j and is defined as

$$w_{ij} = \frac{e^{\frac{|\mathbf{x_i}-\mathbf{y_j}|^2}{2k^2}}}{\sum_p e^{\frac{|\mathbf{x_i}-\mathbf{y_p}|^2}{2k^2}}}$$

so that an energy function E can be defined:

$$E = -\alpha k\sum_i \log(\sum_j \frac{e^{\frac{|\mathbf{x_i}-\mathbf{y_j}|^2}{2k^2}}}{\sum_p e^{\frac{|\mathbf{x_i}-\mathbf{y_j}|^2}{2k^2}}}) + \frac{\beta}{2}\sum_j |\mathbf{y_{i+1}} - \mathbf{y_i}|^2$$

with $\Delta\mathbf{y_j} = -k\frac{\partial E}{\partial \mathbf{y_j}}$. The free parameter k is slowly decreased along the procedure, in a similar way as what is done in simulated annealing algorithms, so that the attraction of the cities becomes more and more specific to their closer path points. The analysis of this energy function was done in [122], for large values of k, the energy has only one minimum but the energy bifurcates as k decreases and then present several local minima. The solution reached by the Elastic Net algorithm is not necessarily the global optimum but the obtained minimum is not too far from the optimal solution [123].

After the original model of Durbin [123], Goodhill et al [124] proposed an implementation of the Elastic Net that generates stripe patterns similar to those observed for ocular dominance domains. In this setting, the cities with positions $\mathbf{x_i}$ represent LGN units scattered on two parallel planes separated by a small gap in 3 dimensional space. The cortex is then an elastic sheet with positions $\mathbf{y_j}$ having the same change rule a explained above. The width of stripes appearing in this model is controlled by the ratio between the distance separating the two LGN sheets and the distance between two neighboring LGN units. This model can test the effect of monocular deprivation by changing the value of α for one of the two eyes. The result obtained in [125] is that the stripes associated with the

eye with the smallest α get lower periodicity without changing the periodicity of the stripes associated to the other eye. Similar principles have been successfully applied to the formation of orientation selectivity.

3.3.2 Pattern formation dynamics.

In the previous model, the formation of V_1 is reduced to the solving of a mapping problem between spaces of different dimensions. The Elastic Net algorithm is successful in reproducing the stripe patterns observed for ocular dominance map and in predicting the effect of visual deprivation but it does so in a rather artificial manner and the history of this pattern formation may be very different from that observed in biology. Pattern formation is a well known phenomena in physics and biology and the possible dynamical systems used to model it are reviewed in [126] and [127]. We will see how these universal models can be used for the formation of ocular dominance stripes and orientation preference maps by integrating physical constrains and symmetries. In these kind of models some biologically relevant aspects are forgotten, for example, the fast dynamic corresponding to the neuronal activity is not taken into account to emphasize the slow dynamics of the synapses.

Ocular dominance stripes The model presented here was first implemented in [128]. It is based on simple rules for the growth of synapses:

- Locally, synapses from the same type activate their growth but inhibit the growth of the same synapses in an annular region around the growth region.
- Synapses of different eyes have opposite influences.

The influence of cells of the same type (w_{RR} or w_{LL}) can then be represented by a difference of Gaussians of the distance r between the two synapses considered:

$$w = Ae^{-\frac{r^2}{d_1}} - Be^{-\frac{r^2}{d_2}}$$

with $d_1 < d_2$ and the influence between synapses from different eyes (w_{RL} or wLR) is represented by the same functions with opposite signs. By this way, the symmetry between the left and the right eye is taken into account. The growth rate of synapses from the right eye is given by:

$$s_R = w_{RR} \star n_R + w_{LR} \star n_L$$

where $\star$ stands for the spatial convolution, with for functions f and g defined on the domain D:

$$f \star g = \int \int_D g(|\mathbf{r} - \mathbf{r}'|) f(\mathbf{r}') d\mathbf{r}'$$

and n_R, n_L are the spatial densities of synapses.

The temporal variations in the number of synapses of both types are also constrained in this model by the fact that the densities should stay in a reasonable range. As densities, it is positive and it is also supposed that at any given point $\mathbf{r}$, the local density is bounded by the maximal density N, $n_R(\mathbf{r}) < N$. The general form of temporal variations in synapses density of the right eye is thus :

$$\frac{\partial n_R(\mathbf{r})}{\partial t} = s_R f(n_R)$$

with f, a function constrained by $f(0) = 0$ and $f(N) = 0$. These constrains can be satisfied by taking for example $f(n_R) = n_R(n_R - N)$ and the dynamical system for the two spatial densities is:

$$\begin{cases} \frac{\partial n_R(\mathbf{r})}{\partial t} & = (w_{RR} \star n_R + w_{LR} n_L) n_R^2 (1 - \frac{N}{n_R}) \\ \frac{\partial n_L(\mathbf{r})}{\partial t} & = (w_{RL} \star n_R + w_{LL} n_L) n_L^2 (1 - \frac{N}{n_L}) \end{cases}$$

The total number of synapses can be considered as locally constant and uniform across the area: $n_R(\mathbf{x}) + n_L(\mathbf{x}) = N$ and the system is then reduced to a single equation:

$$\frac{\partial n_R}{\partial t} = (w_{RR} \star (2n_R - N)) f(n).$$

There are 3 fixed points to this equation:

- $n = 0$ is stable because at first order

$$\frac{\partial n}{\partial t} \approx -Nnw \star N < 0$$

.

- $n = N$ is also stable for symmetry reasons so that if all synapses becomes wired to an eye at time T, there won't be any synapses from the other eye after this time.
- $n = \frac{N}{2}$ is then unstable.

To consider how instabilities develop near the unstable state, we consider a small perturbation $x = n - \frac{N}{2}$. The evolution of such a perturbation is driven by:

$$\frac{\partial x}{\partial t} = \frac{N^2 (w \star x)}{2}.$$

To analyze the transient behavior of the perturbation, it is decomposed as follow, $x = x_0(\mathbf{r}) e^{\lambda t}$. The evolution equation becomes independent of the time variable:

$$\lambda x_0(\mathbf{r}) = \frac{N^2 (w \star x_0(\mathbf{r}))}{2}.$$

The Fourier transform of this equation gives a dispersion relation indicating the stable and unstable modes for the propagation of instabilities (Note that convolution becomes a product after Fourrier transform).

$$\lambda(\mathbf{k}) \hat{x}_0(\mathbf{k}) = \frac{N^2 W(\mathbf{k}) \hat{x}_0(\mathbf{k})}{2}.$$

Unstable modes grow when $\lambda(\mathbf{k}) > 0$ that is when

$$W(\mathbf{k}) = \int\int_D w(|\mathbf{r} - \mathbf{r}'|)e^{i\mathbf{k}.\mathbf{r}'}d\mathbf{r}' > 0.$$

The Fousrier transform of a Gaussian function is a Gaussian function so that:

$$W(\mathbf{k}) = 2\pi \int_0^\infty (Ae^{-\frac{|r-r'|^2}{d_1}} - Be^{-\frac{|r-r'|^2}{d_2}})e^{ikr'}dr'$$

$$W(\mathbf{k}) = A\sqrt{\pi d_1}e^{-\pi^2 d_1 k^2} - B\sqrt{\pi d_2}e^{-\pi^2 d_2 k^2}$$

. The periodicity of ocular dominance stripes will be given by $k^\star$ at which W is maximum, around 0.45 mm with the parameters indicated in fig 3.9.

In order to model the non-uniform distribution of biological markers, like cytochrome oxydase blobs, the upper bound on the density of synapses can be taken as spatially periodic function, $N(r) = \bar{N} + \kappa u(r)$. It was shown in [129] that the pattern formed in this modified model get aligned with the synaptic density.

Orientation preference maps. The formation of orientation columns can be treated in a similar way to that of ocular dominance stripes even if there are some differences. The ocular dominance at a point of the cortex was quantified by a single number n but the orientation preference and selectivity of a column is measured by a vector or its complex representation z. The argument of z is related to the preferred orientation θ of the column:

$$\theta = \frac{1}{2}\arctan\frac{\Re(z)}{\Im(z)}$$

with the factor $\frac{1}{2}$ restricting θ between $-\frac{\pi}{2}$ and $\frac{\pi}{2}$. The module $|z|$ quantifies the selectivity of the response of the column to a bar of orientation θ, it is related to the amount of cells in the column having θ as preferred orientation.

Orientation preference maps are organized around pinwheel singularities where all orientations collapse. Depending on whether the orientations turn clockwise or anticlockwise around the singularity, a positive or negative topological charge is attributed to the singularity. This topological charge can be calculated from the orientation preference field:

$$\mathcal{Q}_A = \frac{1}{2\pi}\oint_C \nabla\theta(\mathbf{r})d\mathbf{s}$$

where the integral is taken over a closed contour surrounding the singularity.

The orientation maps can be characterized by their spatial correlation functions between points at positions $\mathbf{r_1}$

$$C(\mathbf{r_1}, \mathbf{r_2}) =< z(\mathbf{r_1})\bar{z}(\mathbf{r_2}) >$$

$$C^\star(\mathbf{r_1}, \mathbf{r_2}) =< z(\mathbf{r_1})z(\mathbf{r_2}) >$$

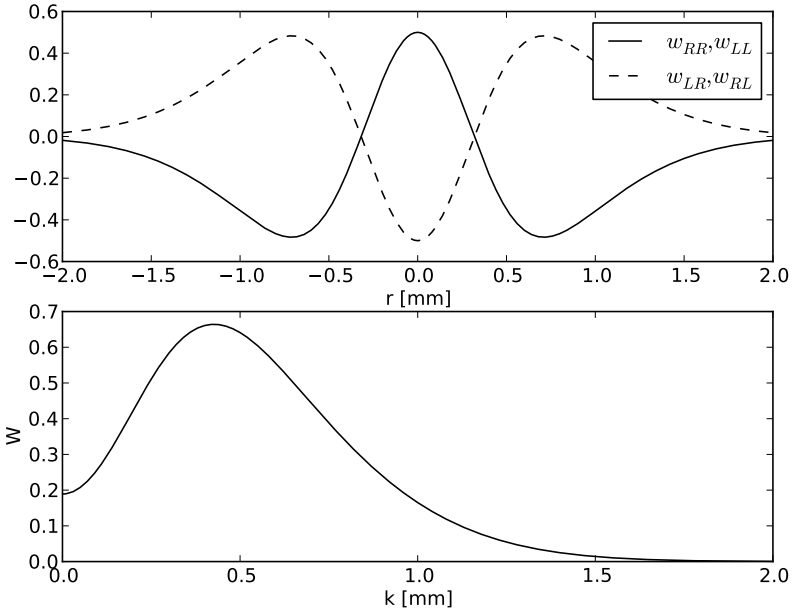


Figure 3.9: **Swindale model for the formation of ocular dominance stripes** - (Top) Kernels describing the influence of synapses from the same eye (w_{RR}, w_{LL}) or from opposite eyes (w_{RL}, w_{LR}). (Bottom) Fourrier transform of the kernel corresponding to synapses from the same eye. The wavelength, $k^\star$, for which it reaches maximum indicates the typical periodicity of the stripes.

The dynamics for the change in z can be written in a general form:

$$\frac{\partial z}{\partial t} = \mathcal{F}[z]$$

Taking a similar approach to the model for the ocular dominance stripes, the following form for $\mathcal{F}$ was proposed in [130]:

$$\mathcal{F}[z] = (w \star z) f(|z|)$$

with w the difference of Gaussians defined in the description of the model for the formation of ocular dominance stripes and $f(|z|) = Z - |z|$ providing an upper bound Z on the selectivity of a column. The only difference with the previous model for the formation of ocular domains is that the field o takes real values whereas for orientation preference, the field z takes complex values and thus similarly to what was shown for ocular dominance, the spatial frequency of the orientation map will be given by $k^\star$ for which the Fourrier transform of w_{RR} is maximal.

The following model of $V1$ was proposed in [131]:

$$\mathcal{F}[z] = \mathcal{L}[z] + \eta$$

without taking an explicit form for the deterministic kernel $\mathcal{L}$ and the noise term η but just constraining the symmetries of $\mathcal{F}$:

- $\mathcal{F}$ is invariant by translation: $\mathcal{F}[T_{\mathbf{R}} z] = T_{\mathbf{R}} \mathcal{L}[z]$ with $T_{\mathbf{R}} z(\mathbf{r}) = z(\mathbf{r} + \mathbf{R})$.
- $\mathcal{F}$ is invariant by rotation: $\mathcal{F}[R_\alpha z] = R_\alpha \mathcal{F}[z]$ with R_α the rotation matrix of angle α: $\begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix}$
- $\mathcal{F}$ is invariant under phase shift: $\mathcal{F}[e^{i\phi} z] = e^{i\phi} \mathcal{F}[z]$.

Phase shift invariance implies that $\mathcal{F}[0] = 0$ and then the homogeneous state $z(\mathbf{r}) = 0$ is a stationary solution of the system. The phase shift invariance also implies that

$$< e^{i\phi} z(\mathbf{r_1}) e^{i\phi} z(\mathbf{r_2}) > = < z(\mathbf{r_1}) z(\mathbf{r_2}) >$$

so that $C^\star$ is null everywhere. The translation invariance implies that $C(\mathbf{r_1}, \mathbf{r_2})$ only depends on the distance $x = |\mathbf{r_1} - \mathbf{r_2}|$. The characteristic wavelength Λ of the orientation map can then be computed by considering the maximum in the Fourrier spectrum of the correlation function:

$$P(k) = \frac{1}{2\pi} \int C(x) e^{ikx} dx$$

Thanks to an analogy with the physics of defects, the analysis of this spectrum gives a lower bound for the density of pinwheels $\rho = \frac{\pi}{\Lambda^2}(1 + \alpha)$ with $\alpha > 0$. This result suggest that for animals with an abnormally low pinwheel density

compared to the wavelength of the orientation map, a phase of pinwheel anhihilation must occur before the map stabilizes. This linear model gives a lower bound on the pinwheel density but it doesn't have any saturation and if it is run without a strong randomness term, the pinwheel density will grow unbounded. A solution to this problem is to consider a non-linear part in $\mathcal{F}$.

The orientation maps are supposed to emerge through Turing instability and an explicit form for $\mathcal{F}$ up to third order can be inspired by pattern formation theory [126]:

$$\mathcal{F} = \mathcal{L} + \mathcal{N}_2 + \mathcal{N}_3$$

with

- $\mathcal{L} = a - (k_c^2 + \nabla^2)^2$, the Swift-Hohenberg operator for describing hydrodynamic instability with k_c the wavelength of the emerging instability.
- $\mathcal{N}_2 = 0$, for symmetry reasons.
- $\mathcal{N}_3[z(\mathbf{r})] = (1-g)|z(\mathbf{r})|^2 z(\mathbf{r}) + (g-2)\int K_\sigma(\mathbf{r}'-\mathbf{r}) z(\mathbf{r})(|z(\mathbf{r}')|^2 + \frac{1}{2}\bar{z}(\mathbf{r}) z(\mathbf{r}')^2) d\mathbf{r}'$ with the kernel $K_\sigma = \frac{1}{2\pi\sigma^2} e^{-\frac{\mathbf{r}^2}{2\sigma^2}}$ accounting for long range interactions and g tuning the fraction of non-linearity coming from local and non-local interactions.

The detailed analysis of this model near criticality, $a << 1$, together with the phase diagram in parameters $(\frac{\sigma}{k_c}, g)$ can be found in [132]. It is shown that for this non-linear model, the pinwheel density will be close to π thus avoiding the unbounded proliferation observed with the purely linear model.

The symmetries considered above are only an approximation of what is observed in the visual cortex. The appropriate invariance is the following: $\mathcal{F}[R_\alpha z] = R_\alpha \mathcal{F}[z e^{i\alpha}]$. A model was proposed recently [133] to take this symmetry into account during development with the following non-linear dynamics for the orientation $\mathbf{s_i} = (s_{ix}, s_{ix})$ of the column i at location $\mathbf{r_i}$:

$$\frac{\partial s_i}{\partial t} = \mathbf{s_i}(1 - |\mathbf{s_i}|^2) + \sum_j [J(r_{ij})\mathbf{s_j} + K(r_{ij}(\mathbf{s_i}.\hat{\mathbf{r}}_{\mathbf{ij}})\hat{\mathbf{r}}_{\mathbf{ij}}],$$

with r_{ij} the distance between columns i and j and $\hat{\mathbf{r}}_{\mathbf{ij}}$ the unitary vector directed by $r_i - r_j$ and j running over all columns. The isotropic coupling J is positive on a disk of radius $\frac{R}{2}$ around the column i and negative on an annular region between $\frac{R}{2}$ and R. The long range term K is taken as constant over the map, it scales the anisotropic input to the column which depends on the colinearity and coaxial alignment between orientations of columns i and j, this term is invariant under joint rotation of the orientations and the grid supporting the columns. In the model with full rotation invariance ($K = 0$), pair annihilation of pinwheels leads to maps with no singularity (rainbow patterns). Adding a non-zero anisotropic term (whether positive or negative) result in stable maps with pinwheels. A similar positive anisotropic term is also included in mean field models of $V1$ dynamics and the action of the special euclidean group on $\mathbf{R}^2 \times \mathbf{S}^1$ under which it is invariant is also called the shift-twist representation of the Euclidean group [134], [135].

3.4 Learning of orientation maps.

In the models described in the two previous sections, biological details of the formation of V_1 were avoided by relying on simplifying hypothesis or universal models which behavior can be studied analytically. Models closer to biology consider the coupling between the dynamics of neuronal activity and the dynamics on the synapses. Changes in the strength of a synapse, if it lasts for few seconds to a minute, is called short-term plasticity and is responsible for synaptic depression and facilitation which are supposed to give a computational advantage for optimal detection and adaptation to changing inputs [136]. Changes can also last for hours or days and this long term plasticity implements learning. The long term potentiation (LTP), increasing the synaptic strength between two coactive neurons, was first described in the rabbit hippocampus [137]. The long term depression is necessary to avoid saturation of all synapses to their maximal strength. In the classical Hebb rule [138], summarized in "cells that fire together, get wired together", the synaptic change during learing is related to the averaged dynamics of the presynaptic and the postsynaptic neurons. In spike timing dependent plasticity [93], synaptic changes depend on the precise timing between the presynaptic and postsynaptic spike, LTP (LTD) is induced if the presynaptic neuron spikes few milliseconds before (after) the postsynaptic neuron. Hebbian learning was widely used in the 80's to implement associative memory through Hopfield network as an example of unsupervised learning. In such network, fixed point attractors of the dynamics can be used to store memories. With this unsupervised learning, after a partial presentation of the input, the network dynamics converge towards the closest memory state. In supervised learning, the network is trained on a set of examples, each example is an input and the desired output to this input and the network change synapses to improve the matching between the desired output and the effective input, for example by backpropagating an error signal [139], and the network can classify then new inputs. Computational models of the development of visual cortex were successfully developed in the 80's. These models, described below, are closer to the biological processes associated to learning than phenomenological models but are also hardly tractable analytically.

3.4.1 The Van der Malsburg model for the formation of columnar organisation.

In a seminal paper of 1973 [140], Christopher Von der Malsburg implemented the first model producing orientation maps. It includes a retinal layer with 19 cells and a cortical layer with 169 excitatory (E) and inhibitory (I) cells disposed on hexagonal grids. The activity H_k represents the firing rate of the neuron k in the cortical layer, driven by a linear differential equation:

$$\frac{dH_k(t)}{dt} = -\alpha_k H_k(t) + \sum_l p_{lk} H_l^\star(t) + \sum_i s_{ik} A_i^\star(t)$$

where $\star$ indicates a thresholded version of the signal to which it is applied.

The fixed recurrent connections are wired according to a Colonnier scheme (1966). The connectivity kernels w from which the p_{kl} are defined have truncated Gaussian shapes of extent σ with $w_{II} = 0$, $\sigma_{EE} = \sigma_{EI}$ and $\sigma_{IE} > \sigma_{EE}$. The plastic synapse s_{ik} representing the strength of connection between the afferent cell i and the excitatory cortical cell k evolves according to a version of the Hebb rule:

> *If there is a coincidence of activity in an afferent fiber i and a cortical E cell k then s_{ik} is increased to $s_{ik} + \Delta s$, Δs being proportional to the signal on the afferent fiber i and to the output signal of the E cell k. Then all the s_{jk} leading to the same cortical cell k are renormalized to keep the sum $\sum_j s_{jk}$ constant.*

A set of nine stimuli was presented to the network, each containing seven active retinal cells and representing a bar of a given orientation. After this training period, each cell fires for a specific orientation and cells are activated in clusters reaching a columnar organization close to what is observed in $V1$, see fig 3.10. The input space is thus mapped on the 2 dimensional surface of the cortex as was done in an abstract way for the elastic net.

3.4.2 Learning rules.

The classical Hebbian learning rule for the synaptic weight from cell i to cell k is the following:

$$w_{ik}(t+1) = w_{ik}(t) + \alpha a_i a_k$$

where a_i and a_k are the pre and postsynaptic activities and α is a free parameter called the learning rate. To avoid unbounded growth of the weights, a normalization scheme should be adopted, it can be substractive or multiplicative. This can be illustrated by considering two binary input neurons of activity I_R and I_L connected with positive synaptic weights w_R and w_L to a cortical neuron of activity $O = w_R I_R + w_L I_L$. With substractive normalization, the dynamics on the weights is:

$$\Delta w_R = I_R O - \frac{1}{2}(I_R O + I_L O)$$

$$\Delta w_L = I_L O - \frac{1}{2}(I_R O + I_L O)$$

replacing O with its value gives

$$\Delta w_R = \frac{1}{2}(w_R I_R^2 + (w_L - w_R) I_R I_L + w_L I_L^2)$$

$$\Delta w_L = \frac{1}{2}(w_R I_R^2 + (w_R - w_L) I_R I_L + w_L I_L^2)$$

then taking averages over input presentations with $< I_R^2 >=< I_L^2 >= 1$ and $< I_R I_L >= C$ gives the continuous time dynamics:

$$\frac{d < w_R >}{dt} = \frac{1}{2}(< w_R > +(< w_L > - < w_R >)C + < w_L >)$$

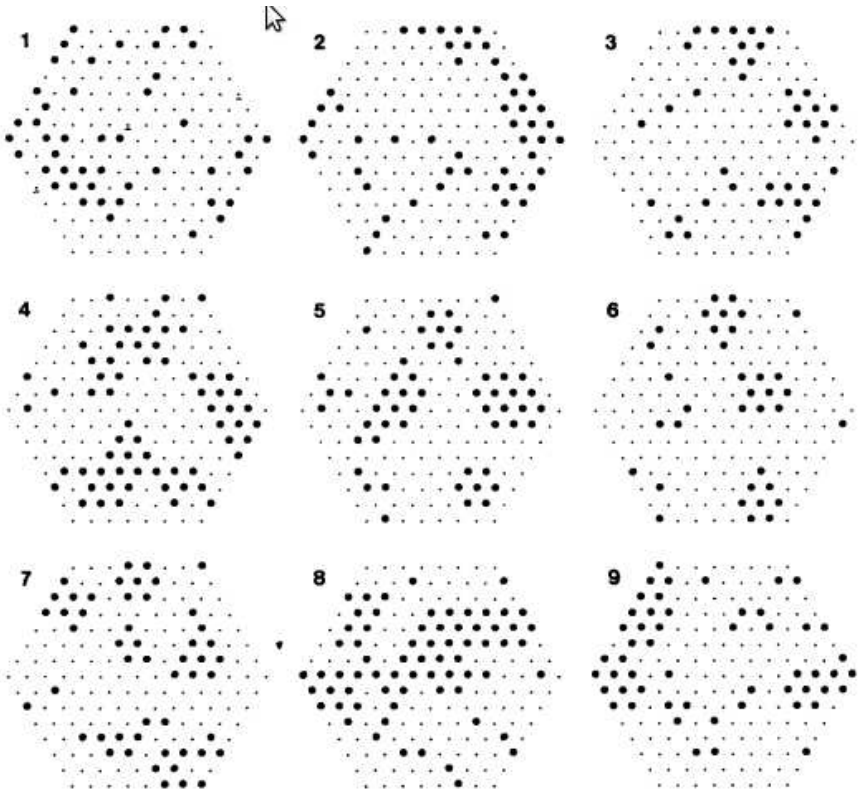


Figure 3.10: **Orientation map.** - Pattern of activity for the 9 oriented bars presented as input (Adapted from [140]).

$$\frac{d < w_L >}{dt} = \frac{1}{2}(< w_L > +(< w_R > - < w_L >)C + < w_R >)$$

which can be written:

$$\frac{d\mathbf{w}}{dt} = \mathbf{A}.\mathbf{w}$$

with $\mathbf{w} = \begin{pmatrix} < w_R > \\ < w_L > \end{pmatrix}$ and $\mathbf{A} = \frac{1}{2}\begin{pmatrix} 1-C & 1+C \\ 1+C & 1-C \end{pmatrix}$. The eigenvalue $\lambda_1 = 1$ is associated to the eigenvector $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and would lead to unbounded growth of the two weights. Weights are usually bounded so that this mode can be ignored. The eigenvalue $\lambda_2 = -C$ is associated to the eigenvector $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$ so that one weight is increasing while the other is decreasing until the system reaches $(0, 1)$ or $(1, 0)$ depending on the stimulation history. If the substractive normalization is used, all weights end up with the value 0 or one and their dynamics is driven by correlations in the inputs. The cortical neurons would then be purely monocular.

To avoid such drastic synapse elimination, the most commonly used normalization scheme is divisive, then for a synapse from cell i to cell k:

$$w_{ik}(t+1) = \frac{w_{ik}(t) + \alpha a_i a_k}{\sqrt{\sum_j (w_{jk}(t) + \alpha a_j a_k)^2}}$$

where j runs across all the input cells of the cell k can be approximated by the local Oja's rule:

$$w_{ik}(t+1) = w_{ik}(t) + \alpha a_i a_k - \alpha w_{ik} a_k^2$$

With such learning dynamics, it was demonstrated that a neuronal network can compute the principal components of the input space [141].

The Bienenstock-Cooper-Munro (BCM) learning rule [142] is another way to avoid the unbounded growth of weights. It provides a mechanism increasing the connection when the sum of weighted input to the neuron is superior to a threshold but also decreasing the weights when the post-synaptic is less than the threshold. The rule in continuous time is:

$$\dot{w}_{ik}(t) = -\epsilon w_{ik}(t) + \phi(a_k(t), \bar{a}_k(t)) a_i(t)$$

with $a_k = \sum_j w_{jk} a_j$ the sum of weighted inputs to the neuron k, $\bar{a}_k(t)$ is the average of the post-synaptic activity over a time T and the function:

$$\phi(a_k(t), \bar{a}_k(t)) < 0 \text{ if } a_k(t) < (\frac{\bar{a}_k}{a_0})^p \bar{a}_k(t)$$

and

$$\phi(a_k(t), \bar{a}_k(t)) > 0 \text{ if } a_k(t) > (\frac{\bar{a}_k}{a_0})^p \bar{a}_k(t).$$

The average post-synaptic activity thus acts as a sliding threshold determining whether the strength of the connection is increased or decreased. This plasticity rule leads to the development of orientation selectivity in a stable manner.

Another set of rules is related to the instrinsic plasticity changing the response function of a neuron rather than the strength of the connection to constrain output activity to have maximum entropy [143] or to follow an exponential distribution [144]. The output y of the neuron is related to its input x by a sigmoidal response function: $y = \frac{1}{1+e^{-(\sigma x+\theta)}}$ with $\sigma > 0$ the slope of the quasi linear part and $\theta > 0$ the threshold. The probability distribution function of y and x are related by $p_y(y) = \frac{p_x(x)}{\partial y/\partial x}$ with

$$\frac{\partial y}{\partial x} = \sigma y(1-y)$$

for a sigmoidal function. A stochastic gradient rule for maximizing a function $L(y)$ is built as follows:

$$\sigma(t+1) = \sigma(t) + \eta \frac{\partial L(y)}{\partial \sigma}$$

$$\theta(t+1) = \theta(t) + \eta \frac{\partial L(y)}{\partial \theta}$$

The differential entropy of the output is $h(y) = -\int_{-\infty}^{\infty} p_y(y) ln(p_y(y)) dy$. With the previous expressions for p_y, $h(y) = \mathbf{E}[\ln(\sigma y(1-y))] - \mathbf{E}[\ln(p_x(x))]$, but the second term doesn't contribute to the output of the neuron so that maximizing the entropy comes to maxmize $L(y) = ln\frac{\partial y}{\partial x}$. We thus obtain the Bell-Sejnowski (BS) rule:

$$\sigma(t+1) = \sigma(t) + \eta(\frac{1}{\sigma(t)} + x(t)(1-y(t))$$

$$\theta(t+1) = \theta(t) + \eta(1-2y(t))$$

with $\eta > 0$, a free parameter similar to the learning rate encountered in synaptic learning rules. The mutual information between the input and the output of the neuron is $I(x,y) = h(y) - h(y|x)$ so that this rule is also maximizing the mutual information between the input and the output. In the rule for σ, correlated activity leads to a decrease of the parameter and for that reason it is another example of anti-Hebbian learning. Constraining the output distribution to be exponential can be formalized as minimizing the Kullback-Leibler divergence between $p_y(y)$ and the target exponential distribution p_{tar} of parameter $\mu > 0$:

$$D(p_y||p_{tar}) = \int_{-\infty}^{\infty} p_y(y) ln \frac{p_y(y)}{\frac{1}{\mu} e^{-\frac{x}{\mu}}}$$

by keeping only terms depending on y, this minimization is equivalent to the maximization of $L(y) = h(y) - \frac{1}{\mu}\mathbf{E}[y]$ that is close to what was done when deriving the BS rule with an additional term for constraining the mean firing rate. The Triesch rule is then

$$\sigma(t+1) = \sigma(t) + \eta(\frac{1}{\sigma(t)} + x(t) - (2+\frac{1}{\mu})x(t)y(t) + \frac{1}{\mu}x(t)y(t)^2)$$

$$\theta(t+1) = \theta(t) + \eta(1 - (2 + \frac{1}{\mu})y(t) + \frac{1}{\mu}y(t)^2).$$

Neuromodulation mediated by acetylcholine or serotonin changes the excitability property of these neurons and could thus support such an intrinsic plasticity. A review of models of neuromodulation is provided in [145]. Hallucination patterns could originate from such changes in intrinsic excitability.

3.4.3 Kohonen network.

A simplified version of the SOM was proposed in [73]. As shortly presented in the first chapter, the algorithm is composed of two parts:

- A competitive stage where the cortical neuron r giving maximal response is selected.
- A cooperative stage where weights are updated locally around the winning neuron.

The rule for updating weights is:

$$w_{ik}(t+1) = w_{ik} + \alpha(a_i - w_{ik})h_{rk}(t)$$

with

$$h_{rk}(t) = e^{-\frac{(x_r - x_k)^2 + (y_r - y_k)^2}{\sigma(t)^2}}$$

the neighborhood function which extent σ is decreased across time in a similar way to what was presented for the elastic net. Because it doesn't include explicitly the activity of the cortical units but only the activity of the retinal units, this model is computationnally less expensive than the original model proposed by Von der Malsburg and it enables mathematical analysis. Kohonen then demonstrated some general properties of its algorithm:

- The input space is more precisely mapped along dimensions having the largest standard deviation.
- Close points in the input space are mapped onto close points in the cortical space.
- Over-represented domains of the input space are mapped onto large domains of the cortical space.

An analysis of the Kohonen network is applied to the formation of primary visual cortex was done in [146]. A cortical unit is located at position $\mathbf{r} = (r_1, r_2)$ in a 2 dimensional space with periodic boundary conditions and is represented by a 5 dimensional feature vector $\mathbf{w}(\mathbf{r})$. The feature vector is a way to code the receptive field properties of the cortical unit and contains:

- x and y, the positions in the retinal space which it is coding for.

- $\mathrm{Re}(z)$ and $\mathrm{Im}(z)$, the components of the vector representing the orientation preference and selectivity of the unit.
- $o = \frac{n_R - N/2}{N}$ characterising the ocular dominance of the unit, with n_R the density of synapses connecting to the right eye and N the density of both types of synapses (right or left). It is 0 if the unit is perfectly binocular, 1 if it receives inputs from the right eye and -1 from the left eye.

The inputs presented to the network are 2 dimensional Gaussians. Each input $\mathbf{a}$ can be represented in the feature space:

$$\mathbf{a} = \begin{pmatrix} x \\ y \\ q\cos(2\phi) \\ q\sin(2\phi) \\ o \end{pmatrix}$$

with (x, y) the position of the center of the Gaussian in the retinal space, q its ellipticity, ϕ the angle between the principal axis of the Gaussian and the horizontal axis in retinal space and o the occularity of the input. Given a probability distribution $P(\mathbf{a})$ of inputs, with $\mathbf{a} \in \mathcal{A}$, the average change is

$$\mathrm{E}(\Delta\mathbf{w}(\mathbf{r})|\mathbf{w}(\mathbf{r})) = \alpha \int_{\mathcal{A}} (\mathbf{a} - \mathbf{w}(\mathbf{r}))h_{\mathbf{s(a)r}}P(\mathbf{a})d\mathbf{a}$$

with the winner unit being selected by:

$$s(\mathbf{a}) = min_{\mathbf{r}}|\mathbf{a} - \mathbf{w}(\mathbf{r})|$$

The stationary states are the ones for which:

$$\mathrm{E}(\Delta\mathbf{w}(\mathbf{r}|\mathbf{w^0}(\mathbf{r})) = 0$$

An obvious stationary state is the purely retinotopic state :

$$\mathbf{w}^0 = \begin{pmatrix} Mr_1 \\ Mr_2 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

where the cortical space is a copy of the retinal space with the magnification factor M. A whole class of stationary states can be obtained by considering translations, reflection or rotations by an angle multiple of π of $\mathbf{w}^0$. The formation of columnar organization associated to orientation and occularity ca, be understood by analyzing the stability through perturbations of these stationary states. This is done in Obermayer 1992 by writing the Fokker-Planck equation in Fourrier space for $\mathbf{u}(\mathbf{r}) = \mathbf{w}(\mathbf{r}) - \mathbf{w}^0(\mathbf{r})$. The order parameters, $(T_i)_{1 \leq i \leq 5}$,

for this equation are the standard deviation of the inputs along each dimension i of the input space:

$$T_i = \frac{\int_{\mathcal{A}}(a_i - < a_i >)^2 P(\mathbf{a})d\mathbf{a}}{\int_A P(\mathbf{a})d\mathbf{a}}$$

with $< a_i >= \int_{\mathcal{A}} a_i P(\mathbf{a})d\mathbf{a}$. The eigenvalues of the linear part of the Fokker Planck equation becomes positive, making the stationary state $\mathbf{w}^0$ unstable, when

$$T_i > T_{\text{thresh}} = \frac{1}{2}\sqrt{e}M\sigma.$$

While inputs keep being similar in their orientation and occularity dimensions, the purely retinotopic state is stable. As the variance in these dimensions crosses a threshold, columnar organization is formed. The fluctuations around the stationary state when inputs are in the suprathreshold threshold regime can be described by the correlation function calculated in [146]:

$$C_{ii}(k) = \frac{\alpha}{2}\pi T_i^2 \sigma^2 \frac{e^{-\frac{\sigma^2}{4}k^2}}{e^{-\frac{\sigma^2}{4}k^2} - \frac{T_i^2 k^2}{M^2}}$$

where $i > 2$ are the indices corresponding to the orientation and occularity dimensions. These fluctuations are finite for $k \to 0$ and grow linearly with the learning rate α.

3.4.4 Plastic lateral connections-LISSOM model

The Kohonen algorithm is based on a winner take all competition so that only one input can be presented at a time. Another pitfall is the symmetry implied by the Gaussian neighborhood function. In visual cortex, the connections between columns are not isotropic and this can be changed by imposing some dynamics on the weights associated to lateral connections.

The Lateral Interactions Synergically Self Organized Maps, LISSOM, model overcomes these problems [147]. The LISSOM can be summarized in five principles:

- The cortical layer is composed by excitatory and inhibitory units disposed regularly on a $N \times N$ grid.
- Inputs from the retina are transmitted through ON and OFF channels of the LGN ($N \times N$). Connections from the retina to the LGN are differences of Gaussian functions, mexican-hat like for the ON channel and reversed for the OFF channel.
- Lateral connections are plastic and initialized as Gaussians with inhibition wider than excitation.
- The activity of a unit is determined by summing linearly the inputs and applying a sigmoidal function or its piecewise linear approximation.

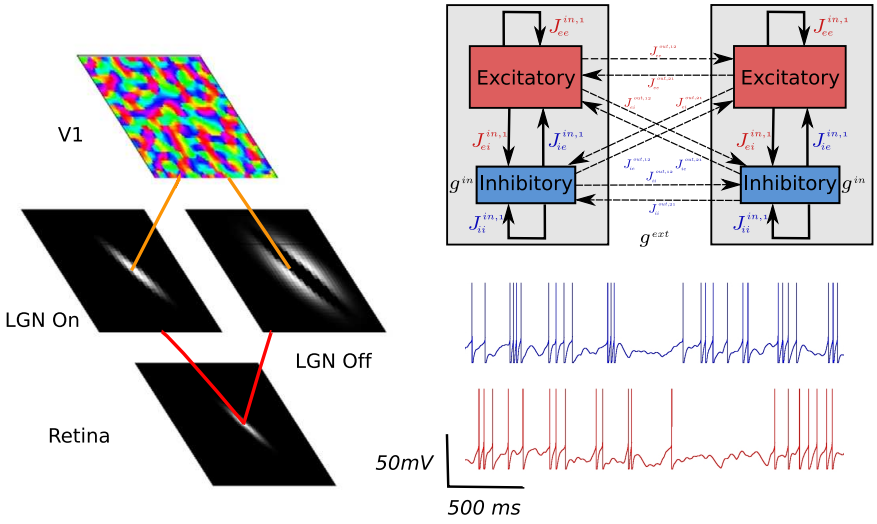


Figure 3.11: **LISSOM Model** - (Left) Description of the LISSOM architecture. Connections in red are fixed and those in orange are plastic. (Right-Top) Wiring diagram between two columns. (Right-Bottom) Typical dynamics of an excitatory (red) and inhibitory (blue) cells.

- Weight dynamics follows Hebbian rule with divisive normalization.

For each stimulus presentation a^{Ret}, the activity of a LGN unit located at position $\mathbf{y}$ is:

$$a^{LGN}(\mathbf{y}) = h(\sum_{\mathbf{x}} w^{Ret-LGN}(\mathbf{x},\mathbf{y})a^{Ret}(\mathbf{x}))$$

where h is the piecewise linear approximation of a sigmoid:

$$h(a) = \begin{cases} 0 & \text{if } a \leq \theta_1 \\ \frac{a-\theta_1}{\theta_2-\theta_1} & \text{if } \theta_1 < a < \theta_2 \\ 1 & \text{if } a \geq \theta_2 \end{cases}$$

The input from LGN to a cortical unit at location $\mathbf{y}$ is

$$s(\mathbf{y}) = \gamma_{LGN}(\sum_{\mathbf{x}\in LGN_{On}} w^{LGN_{On}}(\mathbf{x},\mathbf{y})a^{LGN}(\mathbf{x}) + \sum_{\mathbf{x}\in LGN_{Off}} w^{LGN_{Off}}(\mathbf{x},\mathbf{y})a^{LGN}(\mathbf{x}))$$

At the first step of an input presentation, the cortical activity in $\mathbf{y}$ is

$$a^{Cort}(\mathbf{y},0) = h(s(\mathbf{y})$$

and it includes lateral inputs in the next time steps

$$a^{Cort}(\mathbf{y},t) = h(s(\mathbf{y}+\gamma_E \sum_{\mathbf{x}\in Cort} w^{E}(\mathbf{x},\mathbf{y})a^{Cort}(\mathbf{x},t-1)+\gamma_I \sum_{\mathbf{x}\in Cort} w^{I}(\mathbf{x},\mathbf{y})a^{Cort}(\mathbf{x},t-1))$$

γ_{LGN}, γ_E and γ_I are the relative strengths of afferent, excitatory and inhibitory connections. The rule for updating the weights coming to a unit located in $\mathbf{y}$ is

$$w'(\mathbf{x},\mathbf{y}) = \frac{w(\mathbf{x},\mathbf{y}) + \alpha a(\mathbf{x})a(\mathbf{y})}{\sum_{\mathbf{z}} w(\mathbf{z},\mathbf{y}) + \alpha a(\mathbf{z})a(\mathbf{y})}$$

3.5 Dynamics of the spiking neurons network before and after LISSOM learning.

In the same spirit to what was done for the pinwheel in section, a network is built from a connectivty scheme obtained by LISSOM learning. The LISSOM network is run through the Topographica library [3]. The connectivity is extracted before and after learning to study the different dynamics arising from the corresponding networks.

3.5.1 LISSOM implementation and spiking neurons network.

LISSOM simulation. The parameters used in the LISSOM algorithm are listed in fig 3.12.

[3] Available at http://topographica.org/

Parameter	Value	Description
Pattern of stimulation		
n_I	2	Number of bars at each presentation.
s_I	0.088388	Size of a bar (width of the small axis of Gaussian).
ar_I	4.66667	Aspect ratio of the bar (long axis)/(small axis).
sep_I	0.595826	Minimum of separation between the centers of 2 bars.
Sheets		
d_{Ret}	28	density of retinal units.
d_{LGN}	56	density of LGN units.
d_{Cort}	56	density of cortical units.
Connections		
Retina – LGN		
σ_{center}	0.07385	Size of the center Gaussian for On and Off receptive fields.
$\sigma_{surround}$	0.29540	Size of the surround Gaussian for On and Off receptive fields.
γ_R	2.33	Strength of afferent connections to LGN.
LGN – $\mathbf{V_1}$		
σ_{LGN}	0.27083	Size of the Gaussian from which the weights of the afferent connectivity are initialized randomly.
γ_{LGN}	1.0	Strength of afferent connections to V_1.
α_{LGN}	0.47949	Learning rate for afferent connections to V_1.
$\mathbf{V_1}$ – $\mathbf{V_1}$		
σ_E	0.10417	Size of the Gaussian from which lateral excitatory connections are initialized.
σ_I	0.22917	Size of the Gaussian from which lateral inhibitory connections are initialized.
γ_E	0.9	Strength of lateral excitatory connections.
γ_I	-0.9	Strength of lateral inhibitory connections.
α_E	2.55528	Learning rate for lateral excitatory connections.
α_I	1.80873	Learning rate for lateral inhibitory connections.

Figure 3.12: Parameters used in the LISSOM model for the learning of connectivity.

During this learning phase, all connections have a delay of 0.05. After each stimulus presentation, weights are let evolving during a time $t_{settle} = 2$. A particular feature of the algorithm is a shrinkage of the excitatory kernel similar to what was described for the Elastic Net. At the 300^{th} learning time step, all excitatory connections further than 2 columns of the central unit are pruned and excitatory weights are restricted to this central region as can be visualized on the fig 3.13. The evolution in the shape of receptive fields are illustrated in fig 3.14 for afferent connections and in fig 3.15 for lateral connections. The rotational symmetry is broken by the learning and this results in the emergence of functional properties of the units for detecting orientations. An example of the resulting orientation map is shown in fig 3.16. With a reverse difference of Gaussians connectivity, only two orientations are selected and organize in a checkboard pattern. With long range excitatory connections, stable orientation maps with pinwheels can emerge only if a second $V1$ layer is added [147].

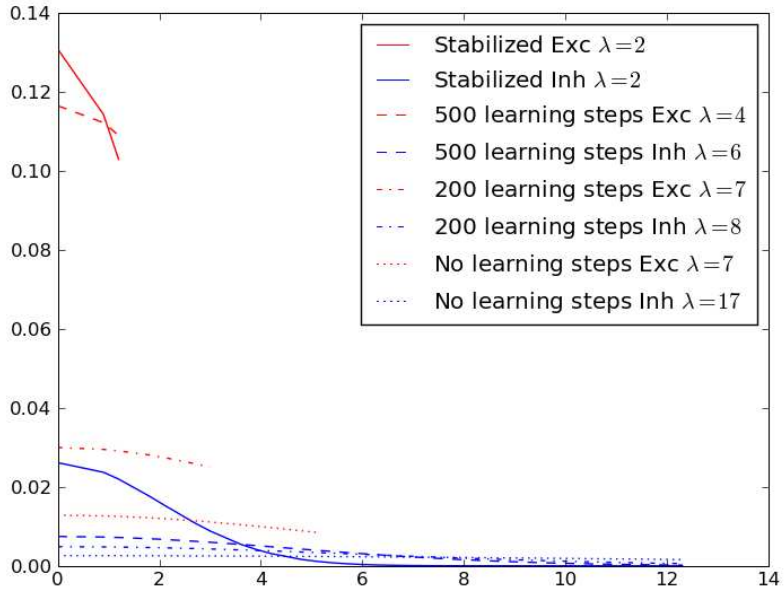


Figure 3.13: **Evolution of the connectivity profile.** - The average profile of connectivity is shown for different stages of the learning with excitatory connections in red and inhibitory connections in blue. Each profile is fitted with a Gaussian function ($e^{\frac{x^2}{\lambda^2}}$) and the corresponding λ is reported in the box.The shrinkage can be detected between the profiles at 200 and 500 learning steps.

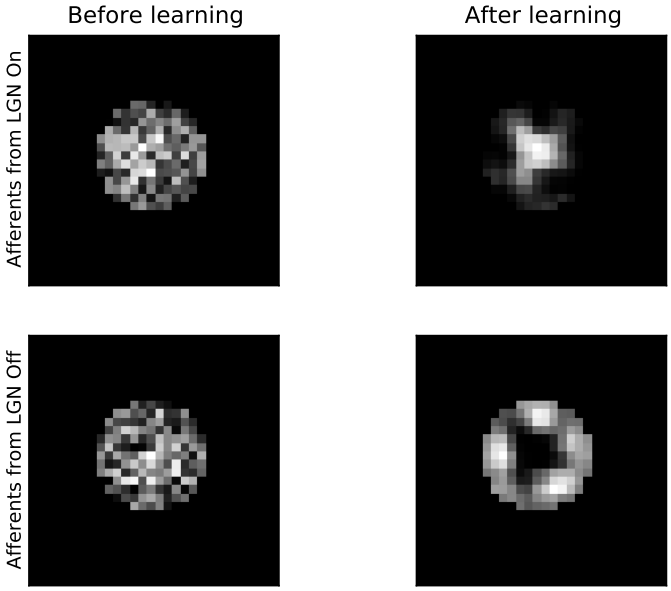


Figure 3.14: **Evolution of the afferent projective fields.** - The projective fields shows the connection weights from a central LGN afferents neuron (ON or OFF) to V_1. Before learning, it is initialized from a 2D Gaussian distribution(Bl.

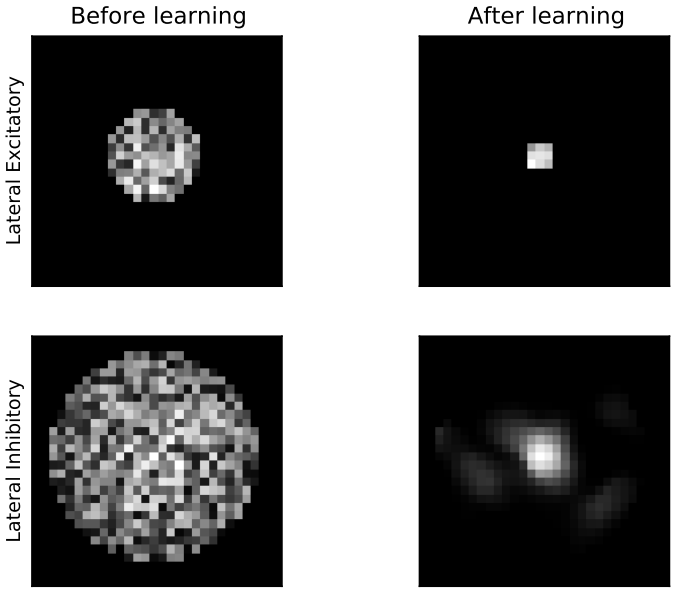


Figure 3.15: **Evolution of the lateral connections.** - The excitatory and inhibitory connections in V_1 are shown before and after learning.

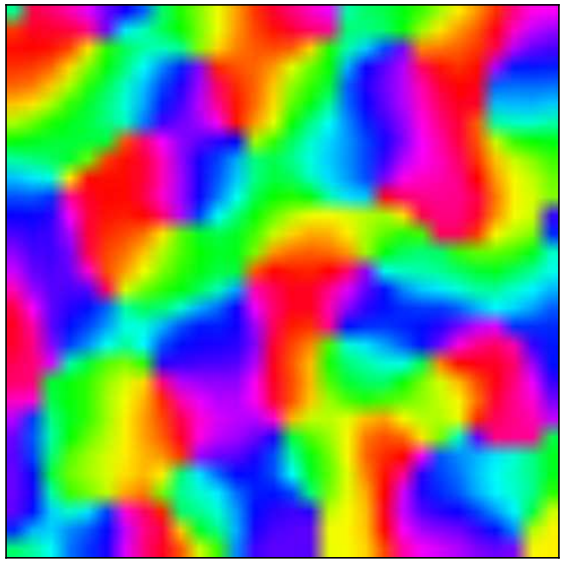


Figure 3.16: **Orientation preference map.** - Resulting map after 20000 time steps. Each color codes for an orientation between $[-\frac{\pi}{2}, \frac{\pi}{2}]$.

Exportation of the connectivity. The V_1 sheet has dimensions 56×56 but in the corresponding spiking neurons network, to avoid boundary effects on the connectivity, only the 30×30 central part is extracted. Weights obtained with LISSOM are between 0 and 1 with many small values, these values can then be used as probability of connection between two columns. This probability will be multiplied by a factor l with a saturation, so that the maximum probability of connection is 1, to scale the influence of lateral connections. The differences between the networks before and after learning can be quantified by measuring their graph properties. The graph of neurons built from the connectivity described above has directed connections with only one connection possible between two neurons. For each neuron, the in degree is the number of neurons connecting to this neuron and the out degree is the number of neurons to which it connects. Considering the same graph as undirected, if there is a connection for each pair in three neurons, there is a triangle passing through these neurons. The clustering coefficient of a node v of the graph is the fraction of possible triangles that actually exist:

$$C_v = \frac{2T(v)}{deg(v)(deg(v)-1)}$$

This is computed using the networkx Python library [4] and shown in fig 3.17. In and out degrees don't change after learning but the clustering coefficient goes from 0.01 before to 0.12 (0.05) for excitatory (inhibitory) neurons after learning. This clustering of the connections is associated to the smaller extent of connectivity profiles shown in fig 3.13 and the break in anisotropy shown in fig 3.15.

Spiking neurons network implementation Implementation of the network is done in with the PyNN implementation of the NEST simulator [148] and analysis of the result is done using Neurotools [5]. The adaptive exponential integrate and fire equations model the dynamics of each neuron with parameters described in chapter II for excitatory and inhibitory cells. Each column is composed of 80 excitatory neurons and 20 inhibitory neurons. Such network are difficult to study because many parameters are involved and insights for the analysis from coarse grained description of the network before analyzing numerical simulations.

3.5.2 Rate description of the $V1$ model before learning.

The connectivity kernels are isotropic at the beginning of the learning process and due to this symetry, the associated rate model can be analyzed easily. A important difference is that in the rate model, the spatial area is supposed to be an infinite plane whereas in the simulations it is a $30x30$ grid. In order to avoid border effects in our simulations, we looped connections to make boundary conditions periodic.

[4] Available at http://networkx.lanl.gov/.
[5] Both available at www.neuralensemble.org

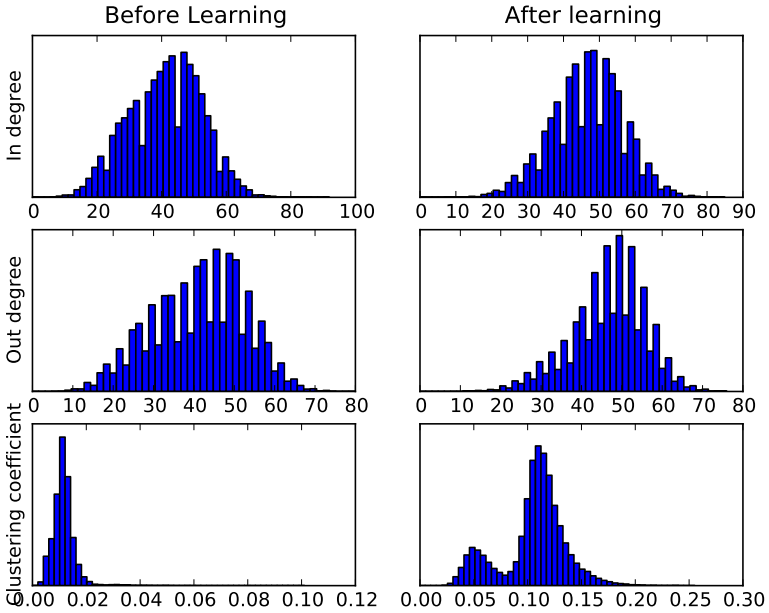


Figure 3.17: **Graph properties of the network** - (Left) Before learning. (Right) After learning.

Eigenforms of the two dimensional model.

Wilson-Cowan equations in continuous space. As we saw in chapter II, the Wilson-Cowan equations give a coarse grained description of the activity of a neuronal network by the dynamics of its average firing rate. The same model can be considered in its spatially continuous version, where the space is represented by a 2D plane with infinite or periodic boundary conditions as was done in [149] [150] [151] [152]. Note that the activity based model described in Chapter 2 can also be extended to this limit. Considering the average rate of a group of neurons located at the position $\mathbf{r}$ in the plane as $E(\mathbf{r},t)$ for the excitatory population and $I(\mathbf{r},t)$ for the inhibitory one, the dynamics evolve as:

$$\frac{\partial E}{\partial t} = -E + f_E(\alpha_{EE}\mu w_{EE} \star E - \alpha_{IE} w_{IE} \star I)$$

$$\frac{\partial I}{\partial t} = -I + f_I(\alpha_{EI}\mu w_{EI} \star E - \alpha_{II} w_{II} \star I).$$

with:

- $\star$ standing for spatial convolution, $f \star g(\mathbf{x}) = \int_{\mathbf{R}^2} f(\mathbf{x}-\mathbf{y})g(\mathbf{y})d\mathbf{y}$
- f_k, with $k \in E, I$, the same kind of response function encountered in Chapter 2, it can be taken as sigmoidal function with $f_k(0) = 0$.
- $w_{kl}(\mathbf{r},\mathbf{r}')$, with $k,l \in E,I$, is the strength of the connection from the group of neurons located in $\mathbf{r}$ to the group of neurons located in $\mathbf{r}'$. This function only depends on the relative distance $|\mathbf{r}-\mathbf{r}'|$ between the two groups. It is positive, bounded with $\int_{\mathbf{R}^2} w_{kl}(\mathbf{r})d\mathbf{r} = 1$ and its Fourrier transform $\hat{w}(\mathbf{k}) = \int_{\mathbf{R}^2} w(\mathbf{r})e^{i\mathbf{kr}}d\mathbf{r}$ is a decreasing function of $|\mathbf{k}|$. A common choice for this connectivity kernels is the Gaussian function $w(\mathbf{r} = e^{-|r|^2}$.
- μ is a parameter modulating the excitability of the network and α_{kl}, $k,l \in E,I$, scales the contribution of population k to the input of population l.

$(E_0(\mathbf{r}), I_0(\mathbf{r})) = (0,0)$ is a stationary solution of the system and we wish to track instabilities which can arise from this uniform state. The stability of the uniform solution can be deduced from the linear system equivalent to the original equations near (E_0, I_0).

$$\frac{\partial E}{\partial t} = -E + f'_E(0)\alpha_{EE}\mu w_{EE} \star E + f'_I(0)\alpha_{IE} w_{IE} \star I$$

$$\frac{\partial I}{\partial t} = -I + f'_E(0)\alpha_{EI}\mu w_{EI} \star E + f'_I(0)\alpha_{II} w_{II} \star I$$

and by the expansions $E(\mathbf{r},t) = \sum_{\mathbf{k}} \hat{E}(\mathbf{k})e^{\lambda t + i\mathbf{kx}}$ and $I(\mathbf{r},t) = \sum_{\mathbf{k}} \hat{I}(\mathbf{k})e^{\lambda t + i\mathbf{kx}}$, we reach the following system for the instability mode $\hat{\mathbf{A}} = (\hat{E}(\mathbf{k}), \hat{I}(\mathbf{k}))$ of wave vector $\mathbf{k}$:

$$\lambda \hat{\mathbf{A}}(\mathbf{k}) = B(\mu)\hat{\mathbf{A}}$$

with

$$B(\mu) = \begin{pmatrix} -1 + f'_E(0)\alpha_{EE}\mu\hat{w}_{EE}(\mathbf{k}) & f'_I(0)\alpha_{IE}\hat{w}_{IE}(\mathbf{k}) \\ f'_E(0)\alpha_{EI}\mu\hat{w}_{EI}(\mathbf{k}) & -1 + f'_I(0)\alpha_{II}\hat{w}_{II}(\mathbf{k}) \end{pmatrix}$$

with $\hat{w}_{kl}(\mathbf{k}) = \int_{\mathcal{D}} w_{kl}(\mathbf{r})e^{i\mathbf{k}.\mathbf{r}}d\mathbf{r}$, the Fourrier transform of the connectivity kernel, which for a Gaussian function is $w_{kl}(\mathbf{k}) = \sqrt{\pi}e^{-\pi^2|\mathbf{k}|^2}$. This eigenvalue problem leads to the dispersion relation $\det(B(\mu) - \lambda I) = 0$. Eigenvalues of B are

$$\lambda_{\pm} = \frac{1}{2}(tr(B) \pm \sqrt{tr(B)^2 - 4det(B)}.$$

If $\lambda_{\pm}$ both have a negative real part, the uniform state is stable. If one of them is positive, unstable modes unfolds resulting in Turing patterns and moreover if $\mathrm{tr}(B) > 0$ and $\mathrm{tr}(B)^2 < 4\det(B)$, those patterns are oscillatory which is known as the Turing-Hopf mechanism for instability [126]. When μ is increased, at least one of the eigenvalues for a given $\mathbf{k}$ becomes positive because $\hat{w}_{kl} > 0$ thus provoking the emergence of patterns of spatial frequency $\mathbf{k}$. The selected $\mathbf{k}$ is the one for which $\lambda(\mathbf{k})$ is maximum that is the one at which $\frac{d\lambda(|\mathbf{k}|)}{d|\mathbf{k}|} = 0$.

Reduced model. The previous model gives the spatial periodicity of the emerging pattern but the rotational and translational symmetries of the connectivity kernels make several doubly periodic patterns possible [6] The relative stability of patterns with different symmetries can be studied on a reduced one dimensional activity based model [7], as proposed in [153]. The activity at a position $\mathbf{r}$ follows:

$$\frac{\partial a(\mathbf{r},t)}{\partial t} = -a(\mathbf{r},t) + \int_{\mathcal{D}} w(|\mathbf{r} - \mathbf{r}'|)f(a(\mathbf{r}'))d\mathbf{r}'.$$

with f a sigmoidal function and w the difference of Gaussians connectivity kernel of extents σ_e, σ_i and amplitudes A_e, A_i as encountered in the Swindale model of part 2 in this chapter:

$$w(|\mathbf{r}|) = \frac{A_e}{\sigma_e}e^{-\frac{x^2}{\sigma_e^2}} - \frac{A_i}{\sigma_i}e^{-\frac{x^2}{\sigma_i^2}}$$

of Fourrier transform:

$$\hat{w}(|\mathbf{k}|) = \sqrt{\pi}(A_e e^{-\pi^2\sigma_e^2|\mathbf{k}|^2} - A_i e^{-\pi^2\sigma_i^2|\mathbf{k}|^2}).$$

The dispersion relation for this simplified model is :

$$\lambda(|\mathbf{k}|) = -1 + \mu\hat{w}(|\hat{k}|)$$

with $\mu = f'(0)$ as a bifurcation parameter. As μ increases there is a range of $k = |\mathbf{k}|$ for which $\lambda(k) > 0$ around the k_0 corresponding to the maximum of λ as ca be seen on fig 3.18.

[6] Periodic in the x and y dimensions.

[7] This model is obtained from the two populations model by considering that the inhibitory dynamics are faster than the excitatory one, replacing the sigmoid in the equation for inhibitory activity by its linear approximation and taking $w_{II} = 0$.

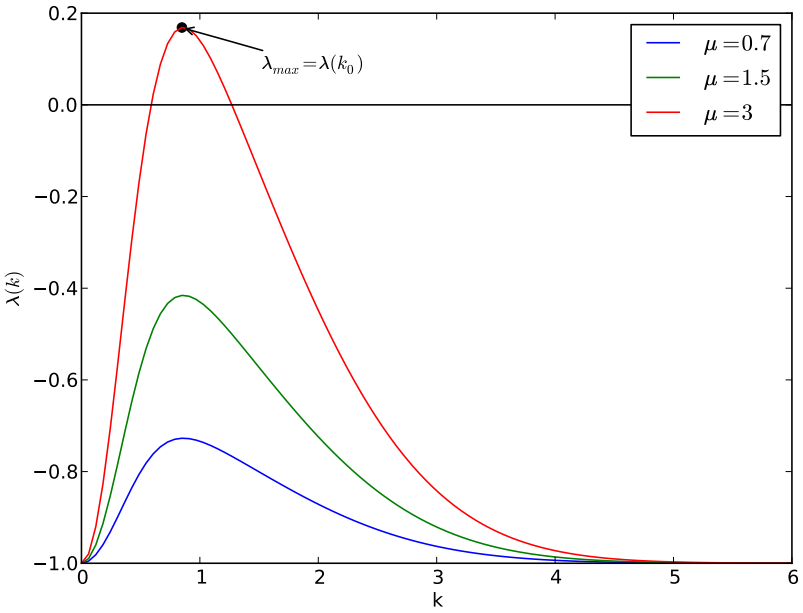


Figure 3.18: **Eigenvalue λ as a function of the norm of the wavevector k.** - This function is plotted for 3 increasing values of μ. When there exist a range over which λ is positive, the k for which λ is maximum indicates the spatial frequency of the emerging pattern.

Near the critical value of μ where a non-uniform pattern arise, the periodicity will be around k_0 but there are several patterns respecting this symmetry:

- For the square pattern, the activity can be expanded as

$$b(\mathbf{x},t) = c_1(t)e^{i\mathbf{k_1 x}} + c_1^*(t)e^{-i\mathbf{k_1 x}} + c_2(t)e^{i\mathbf{k_2 x}} + c_2^*(t)e^{-i\mathbf{k_2 x}}$$

  with $\mathbf{k}_1 = k_0(1,0)$ and $\mathbf{k}_2 = k_0(0,1)$.

- For the rhombic pattern, the activity can be expanded in a similar way but with $\mathbf{k}_2 = k_0(cos\phi, sin\phi)$.

- For the hexagonal pattern, the activity is expanded in the three terms corresponding to the three vectors generating the pattern:

$$\mathbf{k}_1 = k_0(1,0),\ \mathbf{k}_2 = k_0(\frac{1}{2}, -\frac{1}{2}) \text{ and } \mathbf{k}_3 = k_0(-\frac{1}{2}, -\frac{1}{2})$$

The equations for the dynamics of the coefficients of this expansion, c_n, can be obtained by applying perturbation methods and give the relative stability of the possible patterns. The resulting approximation is the following cubic equation for the square or rhombic pattern:

$$\frac{\partial c_n}{\partial \tau} = c_n(\mu - \mu_c - \Gamma_0|c_n|^2 - 2\Gamma_\phi|c_m|^2)$$

with $m, n = 1, 2$ and $m \neq n$. A model including anisotropic long range connections was studied in [134] [154].

Localized bump of activity and traveling wave.

Bump solution. Other types of solutions such as localized bumps and traveling waves also exist in such neural field models (see [155] for a review). An bump solution has been found in [156] when a Heaviside function [8] is used for f in the following one-dimensional layer:

$$\frac{\partial a(x,t)}{\partial t} = -a(x,t) + \int_{-\infty}^{\infty} w(|x-y|)f(a(y,t))dy$$

with w being a difference of Gaussians. A bump solution $A(x)$ of length d and centered on x_0 is defined as a stationary solution such that $A(x) = 0$ on $]-\infty, x_0 - \frac{d}{2}]$ and $A(x) > 0$ on $]x_0 - \frac{d}{2}, x_0 + \frac{d}{2}[$. For such a solution centered on $x_0 = 0$:

$$-A(x) + \int_{R[A(x)]} w(x-x')dx = 0$$

[8] The Heaviside function is defined as follows:

$$H_\theta(x) = \begin{cases} 1 & \text{if } x > \theta \\ 0 & \text{else} \end{cases}$$

with $R[A] = x|A(x) > 0$. By considering this equation at $x = -\frac{d}{2}$, we obtain $h + \int_0^d w(x)dx = h + W(d) = 0$ and this condition is also sufficient. Moreover the stability of this solution is studied by considering that the excited region have moving boundaries $x_1(t) < x_2(t)$ with the slope of the pattern at boundaries being $c_i = \frac{\partial a(x_i,t)}{\partial x}$ $(i = 1, 2)$. From the definition of the excited region, $a(x_i, t) = 0$, $a(x_i + dx_i, t + dt) = 0$ and thus

$$\frac{\partial a(x_i,t)}{\partial t}dt + \frac{\partial a(x_i,t)}{\partial x}dx = 0$$

$$\frac{dx_i}{dt} = -\frac{\partial a(x_i,t)/\partial t}{\partial a(x_i,t)/\partial x}$$

and thus

$$\frac{dx_i}{dt} = \frac{1}{c}(\pm W(x_2 - x_1) + h)$$

so that the evolution equation for the length of the excited region $l(t) = x_2(t) - x_1(t)$ is:

$$\frac{dl}{dt} = (\frac{1}{c_1} + \frac{1}{c_2})[W(l) + h]$$

. The stationary solution with length l_0 is stable if $\frac{dW(l)}{dl}|_{l=l_0} < 0$.

Travelling wave. A traveling wave for this model is a solution $a(z)$, $z = x - ct$ where c is a constant wavespeed, with:

- a monotonic.
- $0 \leq a \leq 1$.
- $a(-\infty) = 0$, $a(\infty) = 1$.

In [157], the existence and uniqueness of such a traveling wave in one dimensional rate model was demonstrated when the response function, f, checks the following properties:

- f continuously differentiable with $f' > 0$.
- $F(a) = -a + f(a)$ has three zeros at 0,$0 < a_0 < 0$ and 1.
- $f'(0) < 0$ and $f'(1) < 1$.

The solution $a(z)$ can be replaced in the model equation, with $z = x - ct$ and $z' = x' - ct$:

$$-c\frac{\partial a(z)}{\partial z} = -a(z) + \int_{-\infty}^{\infty} w(z - z')f(a(z'))dz'$$

k is such that

$$\int_{-\infty}^{\infty} k(x)dx = 1$$

and then

$$-ca'^2 f'(u) = F(a)f'(a)a' + \int_{-\infty}^{\infty} w(z-z')(f(a(z')) - f(a(z')))f'(a(z))a'(z)dz'$$

Integrating according to z gives:

$$-c\int_{-\infty}^{\infty} a'^2 f'(a)dz = \int_0^1 F(a)f'(a)da + \int_{-\infty}^{\infty} w(z-z')(f(a(z')-f(a(z))f'(a(z))a'(z)dzdz'.$$

The last integral can be rewritten:

$$\frac{1}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} w(z-z')(f(a(z')-f(a))(f'(a(z))a'(z) - f'(a(z'))a'(z'))dzdz'$$

then by changing $z' \to z - s$, this integral is:

$$\frac{1}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} w(s)(f(a(z-s)-f(z))(f'(a(z))a'(z) - f'(a(z-s))a'(z-s))dzds$$

which when integrated along z is 0 so that:

$$-c\int_{-\infty}^{\infty} a'^2 f'(a)dz = \int_0^1 F(a)f'(a)da.$$

Moreover

$$\int_0^1 F(a)(f'(a)-1)da = \frac{1}{2}[F(a)^2]_0^1 = 0$$

and then

$$c = -\frac{\int_0^1 F(a)da}{\int_{-\infty}^{\infty} a'^2 f'(a)da}$$

the speed of the wave thus being of the same sign as $-\int_0^1 F(a)da$. In models taking adaptation into account, dynamic patterns are found in [158] [159] [160].

3.5.3 Dynamics in the network of spiking neurons before learning

Periodic boundary conditions: Phase diagram and emergence of static patterns. As was done for the Ice Cubes model, we provide a phase diagram (g_I, g_E) on fig 3.19 indicating the mean and variance of firing rate and the mean of the coefficient of variation of the ISI. There are 3 regimes:

- When excitation dominates, neurons in the whole network fire at maximal frequency. Those regimes are located in the red area of the top panel in fig 3.19.

- When inhibition dominates, neurons of the network fire in an asynchronous irregular fashion. Those regime with high CV are located in the red area of the middle panel in fig 3.19.
- In the well-balanced regime, we see patterns with some parts of the network firing at a high frequency and other parts having a low firing frequency. Those patterns are similar to the eigenforms found in the previously described Turing mechanism. A snapshot of such a pattern is shown in fig 3.20.

The bifurcation diagram for the emergence of the pattern as g_E increases is shown in fig 3.21 for $g_I = 1.5$. At low g_E, the firing rate map is homogeneous with low firing rate. At high values of g_e, the map is also homogeneous but with high firing rate. For intermediate values of g_E, there are some non-homogeneous patterns. Note that, if we only look at the average firing over the network, this transition is not seen.

Finite square boundaries: traveling wave. In another set of simulations, we take keep finite square boundaries and we provide only a local stimulation by connecting Poissonian spike trains as input to the 4 central columns. The firing rate model suggest that the response of the network should be a static bump but due to inhomogeneities of the graph and in the input stimulation, this bump starts moving in a given direction and it is then reflected on the boundaries. For some regimes of (g_E, g_I), the initial bump splits into three bumps rotating on the map.

3.5.4 Dynamics in the network of spiking neurons after learning.

Phase diagram The first parameter varied when building the phase diagram is the balance between maximal conductance of inhibition and excitation $\frac{g_I}{g_E}$ with keeping $g_E + g_I = 0.024nS$ (verifier 0.024) and the second parameter is the lateral strength that is a factor multiplying the probabilities of connection obtained after the LISSOM learning in the rate model. As for the previous model, we detect three regimes:

- The saturated regime (S) is found for low values of $\frac{g_I}{g_E}$ and high values of the lateral strength. In this regime, all neurons fire spikes regularly at their maximal frequency such that the coefficient of variation of interspikes intervals is 0.
- An asynchronous irregular (A) regime ($CV \approx 1$) is found at high values of $\frac{g_I}{g_E}$ and low values of the lateral strength. This well-mixed regime is a good candidate for the on-going activity in awake state.
- A synchronous bursty (B) regime is found in between the two previously described regimes. In this regime, neurons fire spikes in a supra-Poissonian fashion ($CV > 1$).

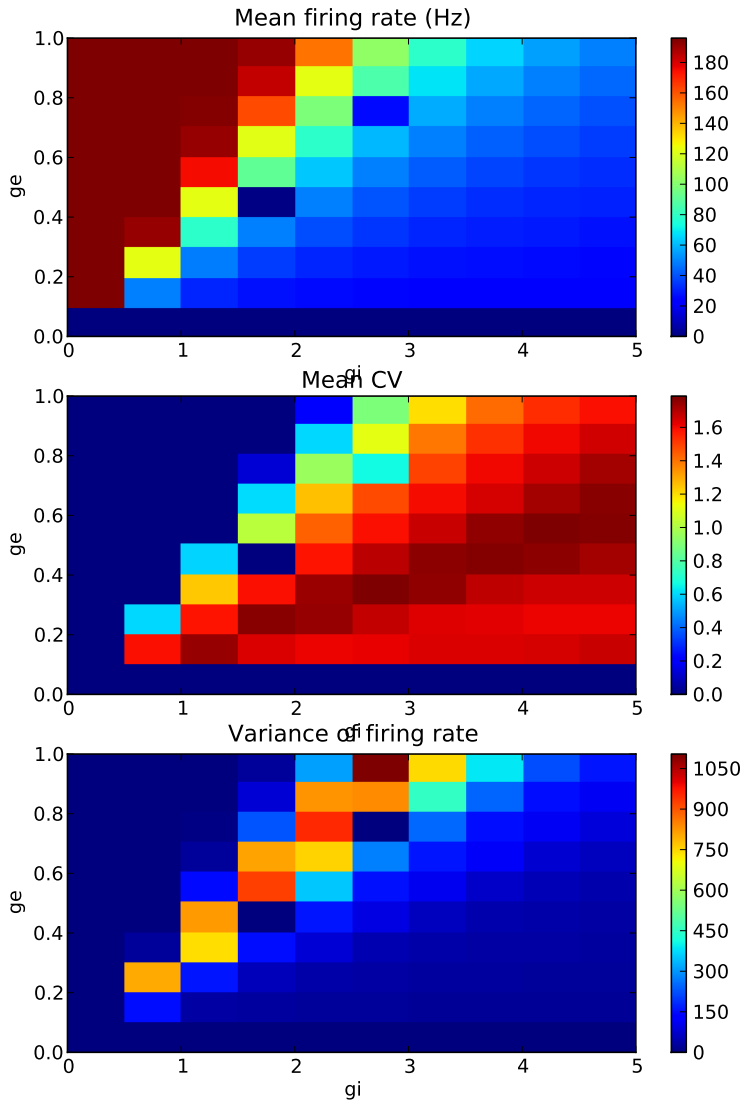


Figure 3.19: **Phase diagrams** (g_E, g_I) **for the network before learning** - The indicated values of conductances g_E, g_I are normalized so that it should be multiplied by $0.006nS$ to get the values used in the simulations.

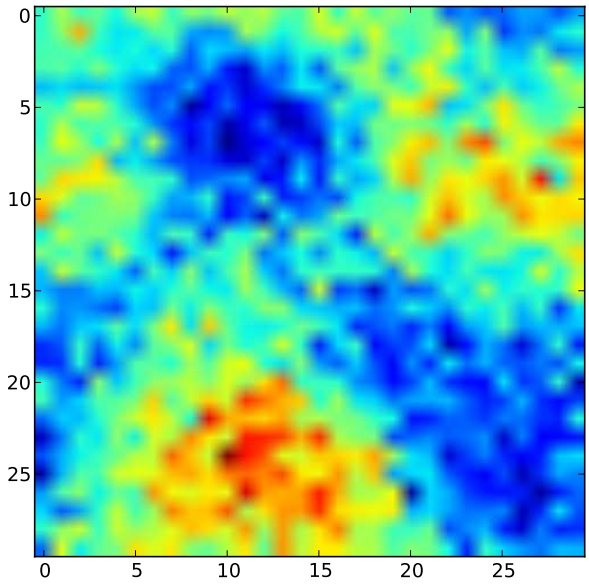


Figure 3.20: **Static pattern** - Firing rate pattern over the 30×30 obtained at the frontier between the excitation dominated regime and the inhibition dominated regime ($g_E = 0.5$,$g_I = 1.5$ in normalized units).

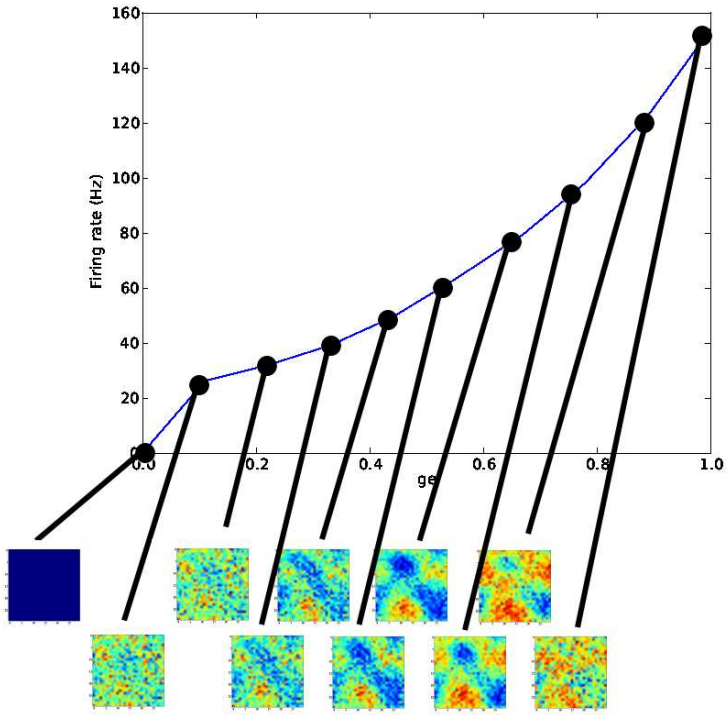


Figure 3.21: **Emergence of a static pattern** - The curve shows the average firing over the net work as a function of g_E with $g_I = 1.5$ in normalized unit and boxes show firing rate maps.

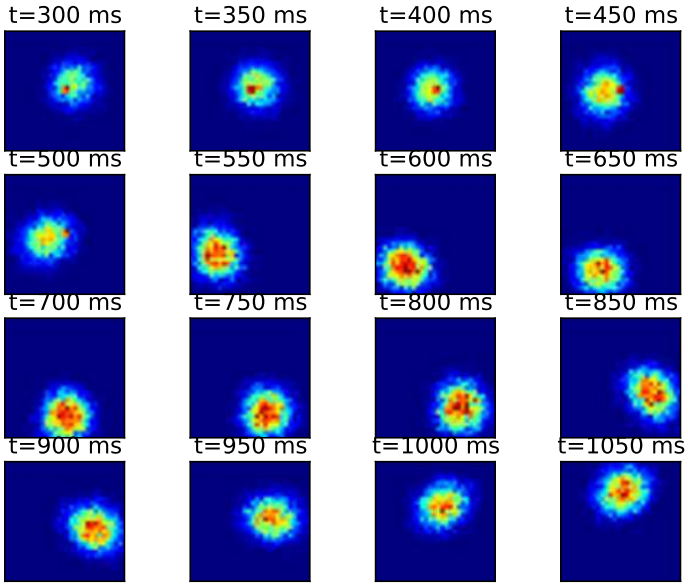


Figure 3.22: **Firing rate map after stimulation of the 4 central columns** - When the stimulation is released (500ms), the bump of activity starts moving across the map.

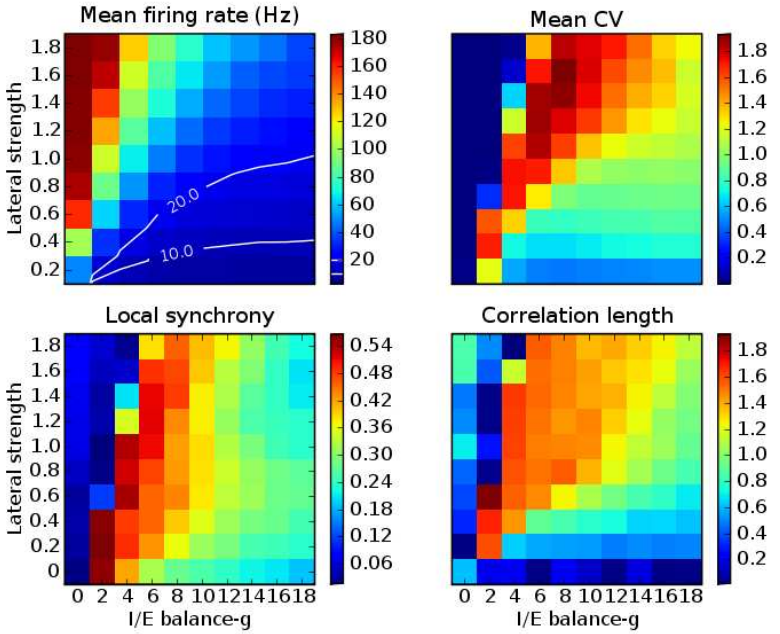


Figure 3.23: **Phase diagram** $(\frac{g_I}{g_E}, ls)$ **for the network of spiking neurons after LISSOM learning** - (Top-left) Average firing rate. (Top-right) Average coefficient of variation of the interspikes intervals. (Bottom-left) Local synchrony c_0. (Bottom-right) Correlation length λ.

Correlations. We calculate the correlation between spike trains as a function of the distance:

$$Cor(d) = \frac{1}{N} \sum_{i,j} Cor(s_i, s_j)$$

where i and j are N random pairs of neurons such that the distance between the columns from which it is extracted is d. The simplest approximation to this function is to assume an exponential decay of correlations as a function of distance:

$$Cor(d) = c_0 e^{-\frac{d}{\lambda}} + c_\infty$$

where c_0 is the local synchrony, that is the correlation averaged over spike trains of the same column. The correlation length λ gives the typical size of patterns emerging in the network. The basal synchrony c_∞ is the correlation between spike trains of columns located very far away from each other. On fig 3.23, it can be checked that the local synchrony and the correlation length have higher values in the busty regime, the correlation length is constant in all that area of the phase diagram.

Description of the activity. In fig 3.24, a phase diagram is shown for parameters (g_{ext}, g_{int}) where g_{int} is the balance $\frac{g_I}{g_E}$ for connections between neurons belonging to the same column and g_{ext} is the balance for connections between neurons belonging to different columns. The sum of conductances $g_E + g_I$ is kept constant ($0.024nS$). The 3 regimes previously described (S, A and B) can also be found in this diagram and we show the activity for a line of columns in regime A and B. In the B regime, activated up states are propagating across the network with slow velocity. The dynamics of a column is represented in fig 3.25 for an example of regime A and an example of regime B. In regime A, spikes seems to occur in an independent fashion and the distribution of the membrane potential is monomodal close to that observed in the main cluster of Chapter 1. In regime B, neurons fire spikes synchronously and the distribution of the average membrane potential is bimodal. The bimodality of cells classified could then result from the network structure when g_E and g_I are tuned in such a regime.

Correlations depend on orientation preference. In the network before learning, the network was invariant under translation and rotation so the correlations between the spike trains of two neurons depended only on the distance separating these two neurons. After learning, the symmetry is broken and the orientation map also have an effect on the correlation structure of the spiking activity in the network. As can be seen on fig 3.26, there is a decay in correlations as a function of distance but also as a function of the difference in preferred orientation. By assuming exponential decay on the distance d and the difference of preferred orientation δ_{phi}:

$$Cor(d, \delta\phi) = c_0 e^{-\frac{d}{\lambda_d} - \frac{\delta\phi}{\lambda_\phi}},$$

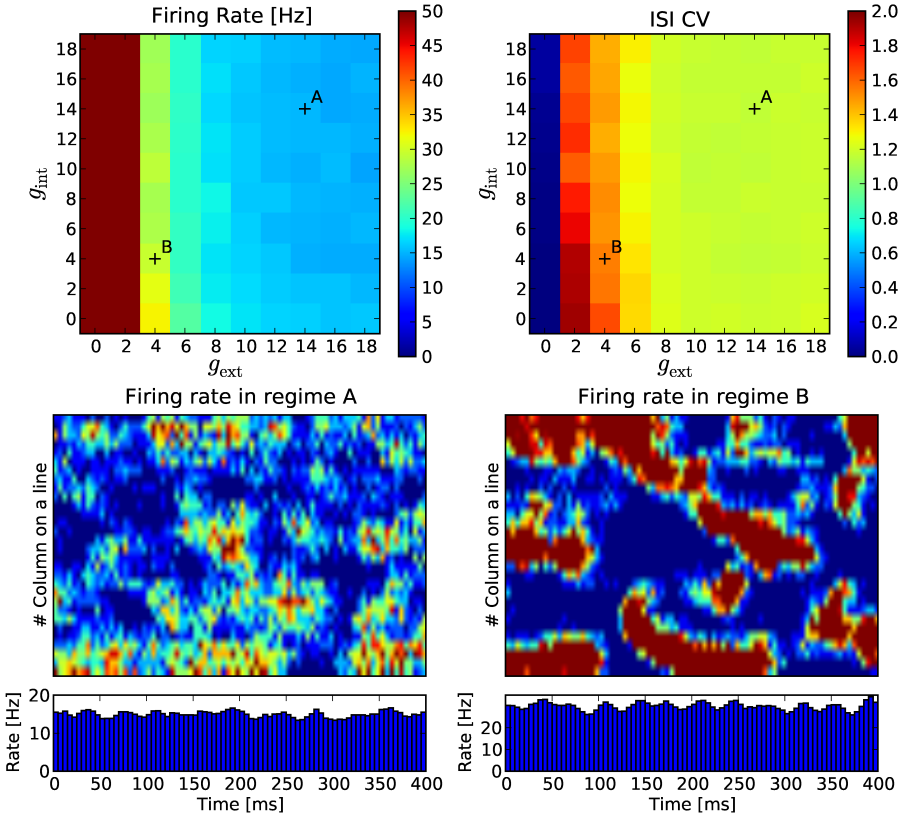


Figure 3.24: **Phase diagram (g_{ext}, g_{int}) and average firing rate on a line of columns** - (Top-left) Average firing rate as a function of (g_{ext}, g_{int}) (rescaled from $g_0 = 0.006$). (Top-right) Average coefficient of variation of the interspikes intervals as a function of (g_{ext}, g_{int}) (Bottom-left) Average firing rate in columns of the central horizontal line of the network over time in the asynchronous regime and average of the firing rate in the network over time. (Bottom-right) Same for regime B.

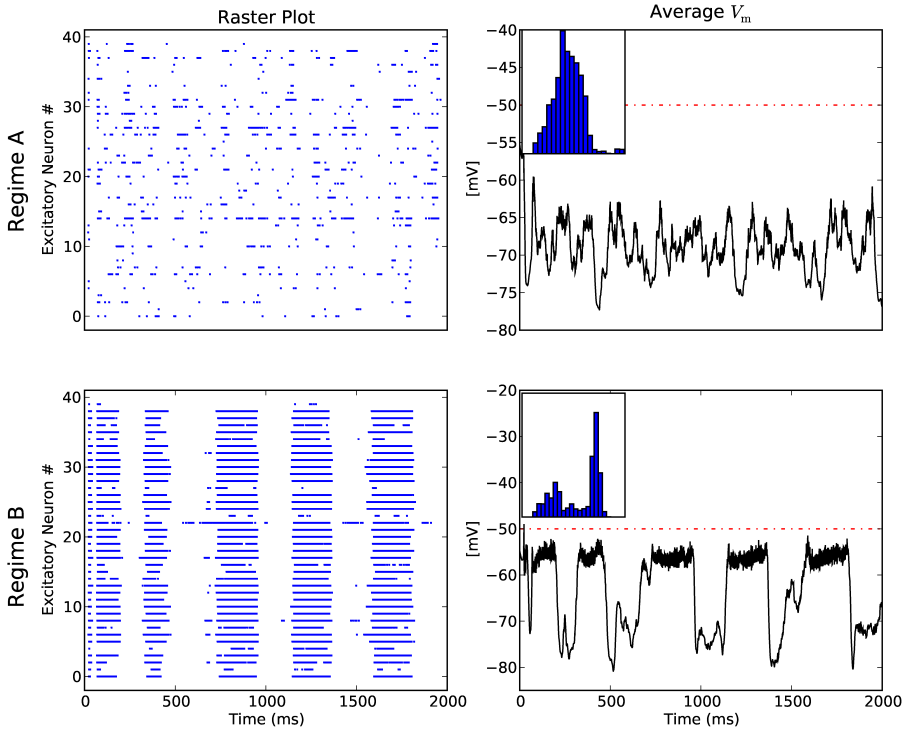


Figure 3.25: **Activity for the asynchronous regime (A) and the bursty regime (B)** - (Left) Spikes raster (each dot represent a spike) for all the neurons in a column of the network. (Right) Temporal dynamics of the average membrane potential and the corresponding distribution (inset).

there are two typical constants characterising the correlation structure of the network, λ_d the spatia correlation length and λ_ϕ the orientation correlation length. The retinotopy and the orientation preference which are reflected in the anatomy are also reflected in the dynamics of the network as was also found in recent LFP recordings [161].

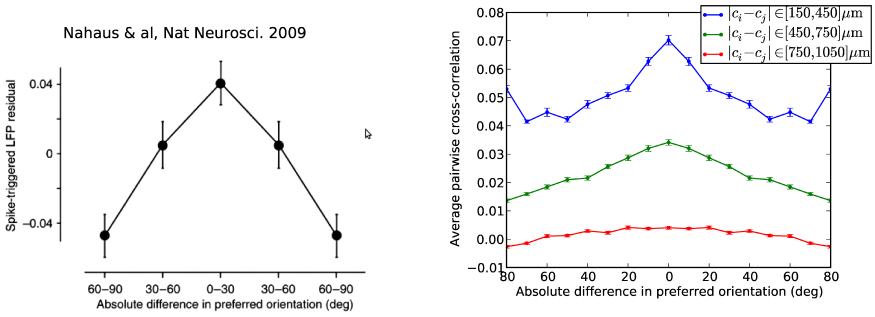


Figure 3.26: **Correlations depending on the difference of orientation preference of the columns.** - (Left) Results from biological experiments [161] showing the cross-correlations in local field potentials depending on the difference of orientation preference. (Right) Average spike trains correlation depending on the difference of preferred orientation $\delta\phi$ for columns separated by a distance d with 3 distances considered. The left part of the curve is a copy of the right part to make the picture similar to the data from [161].

3.5.5 On-going activity in a model of V_1.

The on-going activity of the V_1 model was analyzed before and after LISSOM learning. In both cases, three regimes are found:

- When excitation dominates, the network is saturated at its maximum level of activity.
- When inhibition dominates, neurons of the network fire spikes in a rather independent fashion with low firing rate, similar to the self-sustained asynchronous irregular regime (Kumar 09).
- For balanced situations, the activity in the network is not homogeneous anymore and static or dynamic instabilities occurs.

Before learning, the connectivity is isotropic and as the excitatory conductance is increased, static patterns emerge. By reducing the model to its firing

rate description, we saw that a Turing mechanism may explain the rise of these patterns. After learning, connectivity becomes anisotropic making the mathematical analysis difficult. Analytical results for the neural field model with mexican-hat connectivity and additional weakly anisotropic excitatory connectivity kernel is available in (Golubitsky) but it is very different from our situation where symmetry is broken mainly in the inhibitory kernel. So although a rigorous proof is not possible for now we suggest that the propagating up states in our model could be explained by a Turing-Hopf mechanism in the corresponding neural field model so that the typical length of patterns we observe should also be related to the characteristic length of the connectivity kernel. By this model, we were able to generate up and down states from network mechanisms which is different from other mechanisms relying on intrinsic properties of the cell [162].

3.6 Conclusion.

As described in this chapter, several approaches can be taken to explain the formation of V1 from abstract cognitive architecture to self organizing models implementing a plasticity rule, each having specific application. The ice cubes model provides a framework for natural computation, physical models have been shown to predict the structural properties of feature maps like the periodicity of ocular dominance domains or the pinwheel density of orientation maps and models including plasticity are useful to analyze the connections and dynamics of the primary visual cortex.

With column based network of spiking neurons, we found that in the excitation dominated regime, neurons fire at high rate and in a regular fashion whereas in the inhibition dominated regime, the network sets in an asynchronous irregular state similar to that described for a single column in chapter 2. In models including long range connections, we found another state for balanced excitation and inhibition where patterns of activity emerge, either static for isotropic connections or dynamic for anisotropic connections resulting from learning. Moreover, the structure of correlations in the network after learning reflects the orientation map which is related to visual experience similarly to [25]. Those patterns occurs through an neuronal analog of the Turing instability for the isotropic network. In the network after learning, the up and down states result from a combination of adaptation and inhibition and it would be interesting to study how the duration of up and down states depends on the adaptation variable. In anatomical studies, long range connections are found to be rather excitatory [163] but a simple inversion of initial connections kernels for excitation and inhibition in the rate model doesn't give rise to orientation maps because only two orientations are selected. Two layers are considered in V1 in [164] with short range excitation and inhibition of similar extent in the first layer and long range excitation in the second layer resulting in an orientation map with long range excitation. Long range excitation could also be mediated through long range inhibition targeting only inhibitory cells.

In the ice cubes model, we saw that the ongoing activity can cancel the response of the network to visual stimulation. We also found that the orientation tuning curve was sharper near the pinwheel center than far from this center consistent with the experimental observation that cells stay tuned to orientation near the pinwheel center [165]. Such properties could also be tested in the network with orientation maps resulting from learning and we suggest that the structured correlations encode priors about the statistics of the visual world and some problems in vision like the inference of 3D structure from a 2D image may be solved using such priors. It has already been shown that the level of ongoing activity before the stimulus presentation is a good predictor of the perception of an ambiguous figure [166].

Conclusion

The ongoing activity of the brain characterizes the state of consciousness. in the awake state it is irregular and asynchronous whereas in slow wave sleep there are 1 Hz collective oscillations seen at the single unit level as transitions between up and down states. Under annesthesia, we found similar dynamics to those known for the awake state and the slow wave sleep but also additional classes like cells having a membrane potential with very few fluctuations resulting from synchronous inputs or cells having micro up or micro down states in their membrane potential. The ongoing activity have a wider range than the visually evoked activity and after visual stimulation, the dynamics is close to the main cluster corresponding to the awake state. Another study from Nicolas Benech at the UNIC demonstrated an influence of the ongoing activity on the response properties, with bistable cells having longer latency. The resting state activity is described in human fMRI studies as networks having infraslow ($< 0.1Hz$) correlated fluctuations and recordings of resting state activity are usually long, so we provided a method to represent the infraslow fluctuations of the EEG in an efficient manner using wavelets. At each time, the signal is compressed into a symbol representing its local frequency content and defining a microstate. The classes of neurodynamics we found are the cellular correlates of those microstates and their definition are useful in the monitoring of anesthesia and in the understanding of patterns of ongoing activity recorded at a whole brain level [167]. We presented a collection of parameters which are used to characterize the firing, the distribution and the power spectrum of a cell and we found that those related to the bimodality of the distribution and to the power in high frequencies are useful to separate the clusters. We showed that measures relying on information theory offer a promising approach to multi channels recordings. Several approaches may be taken to classify data and we presented K means clustering, agglomerative tree building and self organized maps. We showed how the partitions resulting from these algorithms can be combined and compared.

Functional neurodynamics, like asynchronous irregular activity of the awake state, fixed point attractors associated to memory storage or limit cycles related to the binding of a coherent percept, are implemented in neuronal networks and can be studied using dynamical systems. To avoid heavy computations and to reduce the parameter space, many properties of neuronal dynamics are studied through phenomenological models which share universal properties, like

the types of bifurcation when a parameter is varied, with the detailed model. Normal form reduction thus gives the simplest polynomial system topologically equivalent to the original system and mean field equations give the evolution of average quantities in the limit of infinite size networks. We reminded how the Hodgkin Huxley model can be reduced to the 2 dimensional FitzHugh Nagumo system and we analyzed local bifurcations of codimension 1 and 2 for this system. At a codimension 3 bifurcation point, a stochastic perturbation via a Brownian motion resulted in complex dynamics mixing several timescales and we suggest that these timescales are related to limit cycles and fixed point attractors lying close to the equilibrium. Interestingly, the first harmonic in the power spectrum had a step-like evolution when the noise variance was increased with the same seed used for the generation of random numbers. A cortical column is modeled as a population of excitatory neurons and a population of inhibitory neurons and their macroscopic activity is described by Wilson-Cowan equations showing multistability and limit cycles. When Hopf bifurcation leads to periodic activity, the dynamics of coupled columns can be reduced to an equivalent phase oscillator and we showed various types of dynamics occuring in a network of phase oscillators from the transition to synchrony as the coupling is increasing to chimera states, in a model with long range connections, where a part of the network is synchronous while the other part is asychronous. In a sparsely connected network, the asynchronous irregular state results from chaotic dynamics when excitation and inhibition are balanced and we provided a Fokker Planck description of the membrane potential in a network of spiking neurons. The effective time scale is very small in the balanced network enabling fast tracking of time varying inputs and can then be used to model attentional effects. Finally, in the article in appendix, we found windows of chaotic behavior in coupled flip-flop oscillators and the ongoing dynamics in a network where cell assemblies are embedded show itinerancy among the fixed points corresponding to stored memories. There are thus wide applications of dynamical systems in neuroscience and complex dynamics modeled in neuronal networks may be used to solve computational tasks [119].

Cortical columns can be coupled together with hard wired or plastic connections to achieve visual function in a model of the primary visual cortex. Several models have been presented to explain how selectivity arises and how the features of the input space are mapped onto the surface of the cortex. In the ice cubes model, standing for a pinwheel in V1, a column have lateral connections only with its nearest neighbours and patterned inputs from the LGN result in simple cell receptive field. In this model, the ongoing dynamics depends on whether excitation or inhibition dominates. When excitation dominates, there is a synchronous regular state with neurons firing at a high rate and no specific response to an oriented bar because the response is lost in the ongoing dynamics. When inhibition dominates, we found an asynchronous irregular state similar to that described for a single column and neurons in a column have a tuning curve reflecting their prefered orientation with this curve being sharper for cells near the pinwheel singularity. Learning was also modeled in a network with long range connections and to fasten simulations, we used a coarse description of

the network with a macroscopic unit representing a column of spiking neurons and weights between two units giving the probability of connection between neurons of the 2 correspondingcolumns. In these models including long range connection, we found structured activity for balanced excitation and inhibition. Before learning, when lateral connection kernels are isotropic, these patterns are static, with some parts of the map firing at a high rate and other parts firing at low rate, and they are related to the eigenforms obtained with a mean field description of the network and considered as a model of hallucinatory perception. After learning using an Hebbian rule on both lateral and feedforward connections to V1 units, the lateral connection kernels become anisotropic, each unit gets selective to orientation and an orientation map emerges. The patterns of the balanced state in the network after learning are dynamic with neurons in each column having collective transitions between up and down states. These transitions are correlated in the network with spiking correlations between neurons of two columns decreasing exponentially as a function of the distance but also as a function of the difference between prefered orientations so that the ongoing activity reflects the visual experience during the learning. Those structured correlation may encode prior knowledge about the statistics of the visual world and used in solving problems related with visual perception. Each perception is thus an inference based on visual stimulation and priors encapsulated in ongoing dynamics.

We thus found a way to characterize several classes of neurodynamics and we found a way to switch from the asynchronous state to structured up and down states by changing the strength of excitation and inhibition in a model of the primary visual cortex. These states are usually associated with specific states of consciousness like slow wave sleep and waking but ongoing activity may also reflect attentional processes and we suggested that ongoing pattern of activity may be useful for cortical computations as reflecting internal knowledge about the world. It would be interesting to provide visual stimulation in the network after learning and to test the network in binocular rivalry, which is a typical exemple where ongoing activity interplay with visual stimulation to give rise to perception [168]. Additional plasticity mechanisms like spike timing dependent plasticity or intrinsic plasticity. The model of the visual cortex can also be used predict the effects of magnetic stimulation in brain computer interface or of a medical drug. The ice cubes model could also be extended to include several pinwheels having long range connections depending on the prefered orientations of connected columns and it would then provide a cognitive architecture for biologically inspired computing. In the column based network we assumed a decoupling of the timescales of the dynamics of neuronal activity and of the synapses but further research is needed to understand interactions between those timescales.

Appendix

Cogn Neurodyn (2009) 3:141–151
DOI 10.1007/s11571-009-9078-0

RESEARCH ARTICLE

Working memory dynamics and spontaneous activity in a flip-flop oscillations network model with a Milnor attractor

David Colliaux · Colin Molter · Yoko Yamaguchi

Received: 30 September 2008 / Revised: 3 February 2009 / Accepted: 3 February 2009 / Published online: 26 February 2009

Abstract Many cognitive tasks require the ability to maintain and manipulate simultaneously several chunks of information. Numerous neurobiological observations have reported that this ability, known as the working memory, is associated with both a slow oscillation (leading to the up and down states) and the presence of the theta rhythm. Furthermore, during resting state, the spontaneous activity of the cortex exhibits exquisite spatiotemporal patterns sharing similar features with the ones observed during specific memory tasks. Here to enlighten neural implication of working memory under these complicated dynamics, we propose a phenomenological network model with biologically plausible neural dynamics and recurrent connections. Each unit embeds an internal oscillation at the theta rhythm which can be triggered during up-state of the membrane potential. As a result, the resting state of a single unit is no longer a classical fixed point attractor but rather the Milnor attractor, and multiple oscillations appear in the dynamics of a coupled system. In conclusion, the interplay between the up and down states and theta rhythm endows high potential in working memory operation associated with complexity in spontaneous activities.

Keywords Working memory · Up down states · Theta rhythm · Chaotic dynamics · Cell assembly

David Colliaux and Colin Molter contributed equally to this work.

D. Colliaux · C. Molter (✉) · Y. Yamaguchi
Lab. for Dynamics of Emergent Intelligence,
RIKEN-BSI, Saitama, Japan
e-mail: cmolter@brain.riken.jp

D. Colliaux
Ecole Polytechnique, CREA, Paris, France

Introduction

During the past 60 years, despite seminal observations suggesting the existence and the importance of complex dynamics in the brain (Nicolis and Tsuda 1985; Skarda and Freeman 1987; Babloyantz and Destexhe 1986), fixed point dynamics has been the predominant regime used to describe brain information processing and more precisely to code associative memories (Amari 1977; Hopfield 1982; Grossberg 1992). More recently, the increasing power of computers and the development of new statistical mathematics demonstrated less equivocally the necessity to rely on more complex dynamics (e.g. Varela et al. 2001; Kenet et al. 2003; Buzsaki and Draguhn 2004). In that view, by extending classical Hopfield networks to encode cyclic attractors, the authors demonstrated that cyclic and chaotic dynamics could encompass several limitations of fixed point dynamics (Molter et al. 2007a, b).

During working memory tasks, tasks requiring the ability to maintain and manipulate simultaneously several chunks of information for central execution (Baddeley 1986), human scalp EEG (Mizuhara and Yamaguchi 2007; Onton et al. 2005) and neural firing in monkeys (Tsujimoto et al. 2003; Rainer et al. 2004) suggested that the theta rhythm (4–8 Hz) plays an important role. The neural basis of the working memory has been widely investigated in primates by using delay to matching tasks. In these tasks, the primate has to retain specific information during a short period of time to guide a forthcoming response. Single cell recordings has shown that during this delay period, some cells located in specific brain areas had increased firing rates (Fuster and Alexander 1971; Rainer et al. 1998).

How these cells can maintain sustained activity is not solved. However, while the coding of information in long-term memory is mediated by synaptic plasticity

(e.g. Whitlock et al. 2006), it seems that working memory and more generally short-term memory relies on different mechanisms, potential candidates being that the information is maintained in the network's dynamics itself (Goldman-Rakic 1995). Recent related observations reported the existence of slow oscillations in the cortex associated with 'flip-flop' transitions between bistable up- and down-states of the membrane potential; Up-states, or depolarized states, being associated with high firing rates of the cells during seconds or more. The transition between up and down states as well as the maintenance of high activity during up states could be the result of network interactions, and the neural basis for working memory (McCormick 2005).

Different models of working memory have been proposed. According to the classification proposed by Durstewitz et al. (2000), one type of model is based on a cellular mechanism of bistability (e.g. Lisman and Idiart 1995). Another type follows to the tradition of classical rate coding scheme proposed by (Amari 1977; Hopfield 1982; Grossberg 1992). In these models (e.g. Brunel et al. 2004; Mongillo et al. 2005), the associative memories are coded by cell assemblies in a recurrent network where the synaptic weights are usually pre-encoded according to a Hebbian rule (Hebb 1949). As a result, the recovery of a memory, i.e. of a cell assembly, leads to persistent activity of that cell assembly through recurrent excitation. This could be described as an up-state, while the attractor to the resting state would be the down-state. Flip-flop transitions between up- and down-states are regulated by applying transient inputs.

Remarkably, these models neglect the presence, and accordingly the possible contribution, of the theta rhythm observed during working memory tasks. However, many physics studies have demonstrated the powerful impact of rhythms and oscillations on synchronization which in turn could play an important role during working memory tasks. Here, to conciliate the cell assembly theory with the presence of brain rhythms, we hypothesize that during up-states the firing rate is temporally modulated by an intrinsic cellular theta rhythm. Synchronization among cellular rhythms leads to dynamical cell assembly formation in agreement with observation of EEG rhythms. Grounding on two previous reports (Colliaux et al. 2007; Molter et al. 2008), we propose here the "flip-flop oscillation network" characterized by two different temporal and spatial scale. First, at the unit scale, a theta oscillation is implemented, second, at the network scale, cell assemblies are implemented in the recurrent connections. We demonstrate how the intrinsic cellular oscillation implemented at the unit level can enrich, at the network level, the dynamics of the flip-flop associative memory.

In the following section, we first formulate the flip-flop oscillation network model. Then, we show that the dynamics for a single cell network has an interesting attractor in term of the Milnor attractor (Milnor 1985). Following that, network dynamics in two coupled system is further analyzed. Finally, associative memory networks are elucidated by focusing on possible temporal coding of working memory.

Model

To realize up- and down-states where up-states are associated with oscillations, two phenomenological models are combined. First, in tradition of Hopfield networks and the theory of associative memories, each cell i is characterized by its membrane potential S_i, and is modulated by the activity of other cells through recurrent connections (Hopfield 1982). Second, to account for the presence of an intrinsic oscillation, each cell is characterized by a phase ϕ_i. The phase follows a simple phase equation defining two stable states: a resting state and a periodic motion (Yamaguchi 2003; Kaneko 2002; Molter et al. 2007c). These two variables are coupled such that, first, an oscillation component $\cos\phi_i$ produces intrinsic oscillation of the membrane potential, second, the evolution of the phase depends on the level of depolarization. As a result, the cell's dynamics results from the non linear coupling between two simpler dynamics having different time constant. Thus, in a network of N units, the state of each cell is defined by $\{S_i, \phi_i\} \in \Re \times [0, 2\pi[\ (i \in [1, N])$ and evolves according to the dynamics:

$$\begin{cases} \frac{dS_i}{dt} = -S_i + \sum_{j=1}^{N} w_{ij} R(S_j) + \Gamma(\phi_i) + I_i \\ \frac{d\phi_i}{dt} = \omega + (\beta - \Lambda(S_i)) \sin\phi_i \end{cases} \quad (1)$$

with w_{ij}, the synaptic weight between cells i and j, $R(S_j)$, the spike density of the cell j, and I_i represents the driving stimulus which enables to selectively activate a cell. In the second equation, ω and β are respectively the frequency and the stabilization coefficient of the internal oscillation.

The spike density is defined by a sigmoid function:

$$R(x) = \frac{1}{2}(\tanh(g(x - 0.5)) + 1), \quad (2)$$

The couplings between the two equations, Γ and Λ appear as follows:

$$\begin{cases} \Gamma(\phi_i) = \sigma(\cos\phi_i - \cos\phi_0) \\ \Lambda(S_i) = \rho S_i \end{cases} \quad (3)$$

where ρ and σ modulates the coupling between the internal oscillation and the membrane potential, and ϕ_0 is the equilibrium phase obtained when all cells are silent ($S_i = 0$); i.e. $\phi_0 = \arcsin(-\omega/\beta)$.

The following set of parameters was used: $\omega = 1$, $\beta = 1.2$ and $g = 10$. Accordingly, $\cos\phi_0 \approx -0.55$. ρ, σ and

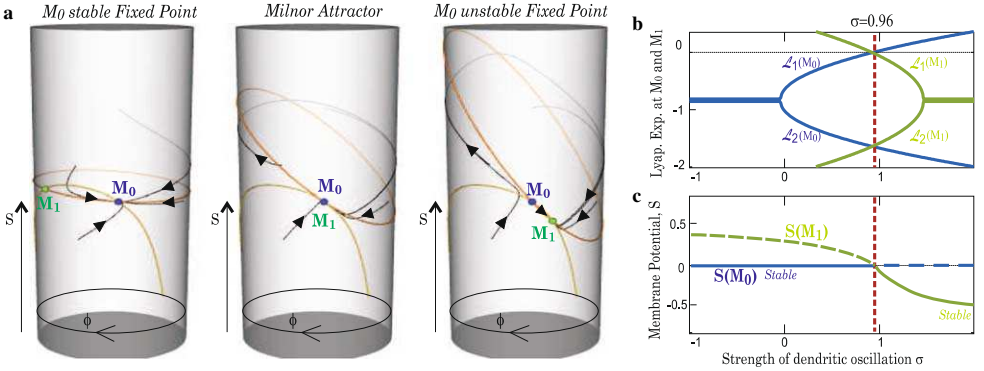


Fig. 1 Fixed points analyses for one cell. **a** Cylinder space (S,ϕ) with nullclines (orange for $dS/dt = 0$, yellow for $d\phi/dt$) and some trajectories. Left to right shows the three possible scenario: M_0 is stable fixed-point for $\mu < \mu_c$, M_0 is the Milnor attractor for $\mu = \mu_c$ and M_0 is unstable fixed-point for $\mu > \mu_c$. **b** Evolution of the two Lyapunov exponents of the system at M_0 (in blue) and M_1 (in green) in function of σ ($\rho = 1$). As expected, one exponent becomes null at $\sigma = \mu_c$. **c** Evolution of the two fixed points, M_0 and M_1, when σ is varied. $\mu = \mu_c$ corresponds to a transcritic bifurcation

w_{ij} are adjusted in each simulations. A C++ Runge–Kutta Gill integration algorithm is used with a time step set to $h = 0.01$ in first simulations and to $h = 0.1$ in Section "Working memory of cell assemblies" simulations (to fasten the computation time). The variation of h leaded to no visible dynamical change. If we consider that $\omega = 1$ represents the 8 Hz theta oscillation, one computational time step represents $0.01/8/2\pi \sim 0.199$ ms in first simulations and ~ 1.99 ms in last section.

One unit

For one unit, the dynamics in Eq. 1 simplifies to:

$$\begin{cases} \frac{dS}{dt} = -S + \sigma(\cos\phi - \cos\phi_0) + I \\ \frac{d\phi}{dt} = \omega + (\beta - \rho S)\sin\phi \end{cases} \quad (4)$$

Next paragraph analyzes the dynamics in absence of any external input ($I = 0$). Then, to understand how information is processed by a neural unit, we observe the dynamics in response to different patterns of stimulation.

Fixed point attractors

When $I = 0$, since $\omega < \beta$, the dynamics defined by Eq. 1 has two fixed points, $M_0 = (0,\phi_0)$ and $M_1 = (S1,\phi_1)$. The linear stability of M_0 is analyzed by developing the Jacobian of the coupled Eq. 4 at M_0:

$$DF|_{M_0} = \begin{pmatrix} -1 & -\sigma\sin(\phi_0) \\ -\rho\sin(\phi_0) & \beta\cos(\phi_0) \end{pmatrix} \quad (5)$$

The eigenvalues are given by:

$$\lambda_{1,2} = \frac{1}{2}\left(\eta - 1 \pm \sqrt{(\eta-1)^2 + 4(\rho\sigma\sin^2\phi_0 + \eta)}\right) \quad (6)$$

where $\eta = \beta\cos\phi_0$. The linear stability of the fixed point M_0 is a function of the internal coupling between S and ϕ, $\mu = \rho\sigma$. With our choice of $\omega = 1$ and $\beta = 1.2$, one eigenvalue becomes positive and M_0 becomes unstable for $\mu > \mu_c = 0.96$. Since μ is the crucial parameter for the stability of the resting state, we fixed here $\rho = 1$ and we analyzed the dynamics according to σ.

Figure 1a shows trajectories and nullclines in the three possible scenario ($\mu < \mu_c$, $\mu = \mu_c$ and $\mu > \mu_c$). The two fixed points M_0 and M_1 (defined by the intersection of the two nullclines) merge for $\mu = \mu_c$. To have a better understanding of the dynamics, Fig. 1b shows the evolution of the Lyapunov exponents at M_0 and M_1 when σ[1] is varied. Figure 1c shows how the two fixed points exchange their stability at criticality. This corresponds to a transcritical bifurcation. Around the bifurcation our two equations (Eq. 1) can be reduced to the normal form of the bifurcation:

$$\begin{cases} \frac{dx_1}{dt} = ax_1^2 + \lambda_1 x_1 \\ \frac{dx_2}{dt} = \lambda_2 x_2 \end{cases} \quad (7)$$

Figure 2 shows how the fixed points changed stability according to the variation of λ_1.

For the fixed point M_0:

[1] Lyapunov exponents aims to quantify the dependency of the dynamics to infinitesimal perturbation. They are computed here analytically as the eigenvalues of the Jacobian (see Ott 1993).

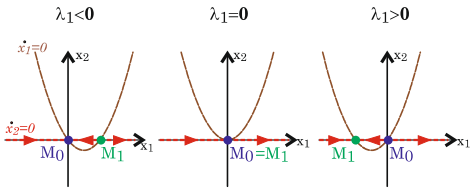


Fig. 2 Normal form reduction near μ_c. The two nullclines $\dot{x_1} = 0$ and $\dot{x_2} = 0$ appear. Plain arrows indicate the direction of the trajectories for $x_2 = 0$

$\mu < \mu_c$: both Lyapunov exponents are negative. All trajectories converge to the trajectories M_0 which is accordingly stable. The other fixed point, M_1, is unstable with $\phi_1 < \phi_0$ and $S_1 > S_0$;
$\mu > \mu_c$: one Lyapunov exponent is negative while the other is positive. It corresponds to the co-existence of both attracting and diverging trajectories. Trajectories attracted to M_0 escape directly from it. M_0 corresponds to an unstable fixed point. M_1 is stable with $\phi_1 > \phi_0$ and $S_1 < S_0$;
$\mu = \mu_c$ one Lyapunov exponent is negative. Accordingly, the basin of attraction of M_0 has a positive measure and many trajectories are attracted to it. However, since the other Lyapunov exponent equals zero, it exists an 'unstable' direction along which the dynamics can escape due to infinitesimal perturbation. Accordingly, M_0 does not attract all trajectories from an open neighborhood and does not define a classical fixed point attractor. However, it is still an attractor if we consider Milnor's extended definition of attractors (Milnor 1985). At M_0, the two nullclines becomes tangent and the two fixed points merge, $M_0 = M_1$.

The particularity and interest of the Milnor attractor is the following. For all values of μ, any initial state of our system is attracted and converges to a fixed point of the system (M_0 or M_1). To escape from that fixed point, and to perform an oscillation, a given amount of energy has to be provided, e.g. as a transient input. The more $\mu \ll \mu_c$ or $\mu \gg \mu_c$, the more the amount of energy required to escape from that fixed point is important. By contrary, at the Milnor attractor (or for $\mu \approx \mu_c$,) the dynamics becomes very receptive and an infinitesimal perturbation can push the dynamics to iterate through an oscillation.

Response to a constant input

Under constant inputs, the dynamics of the system 4 can either converge to a fixed point, either converge to a limit cycle. To obtain a fixed point (S_*, ϕ_*), $d\phi/dt$ must be equal to zero in Eq. 4, which requires:

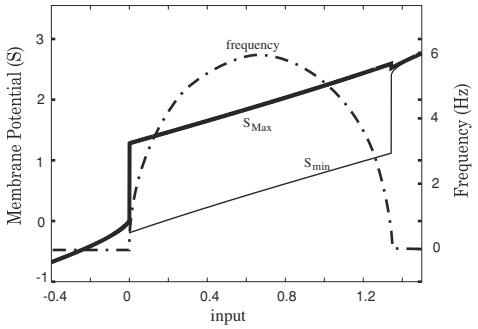


Fig. 3 Evolution of the maximum and minimum values of S, and of the dominant frequency obtained by FFT when the input current is varied

$$\left|\frac{\omega}{\beta - S_*}\right| < 1 \tag{8}$$

and Eq. 4 becomes:

$$\begin{aligned} S_* &= \sigma(\cos\phi_* - \cos\phi_0) + I \\ 0 &= \omega + (\beta - \sigma(\cos\phi_* - \cos\phi_0) - I)\sin\phi_* \end{aligned} \tag{9}$$

Figures 3 and 4a show the evolution of the dynamics for various levels of input. Fixed points are obtained either for small values of I ($S < (\beta - \omega)$ in Condition 8), either for large values of I ($S > \beta$ in Condition 8). In between, oscillatory dynamics occurs. The period of the oscillation can be approximated by identifying S with its temporal average, $\overline{S}$ and by solving the integral $\int_0^{2\pi} \frac{d\phi}{\omega + (\beta - \overline{S})sin(\phi)}$. This approximation gives an oscillation at frequency $\omega' = \sqrt{\omega^2 - (\beta - \overline{S})^2}$, which is in good agreement with computer experiments (Fig. 3).

In Fig. 4a, the dynamics of our system of two equations is compared with the dynamics obtained for $\sigma = 0$ (i.e. for a simple integrating unit as in the rate model) and for the current output from the phase model ($\cos\phi$). As expected from Fig. 3, the coupling of the phase adds oscillation in a large range of inputs, and, as a first order approximation, our unit's dynamics appears as a non-linear combination of both dynamics.

Response to oscillatory input

Since the output of a unit can become the input to other units and can contain oscillations, it is interesting to see how a unit reacts to oscillatory inputs. Figure 4b shows the dynamics obtained when considering an oscillating currents having an increasing frequency. The input current follows the ZAP function, as proposed in (Muresan and Savin 2007): $I = \sin(\alpha t^{\beta})$, with $\alpha = 2\pi 10^{-5}$ and $\beta = 3$.

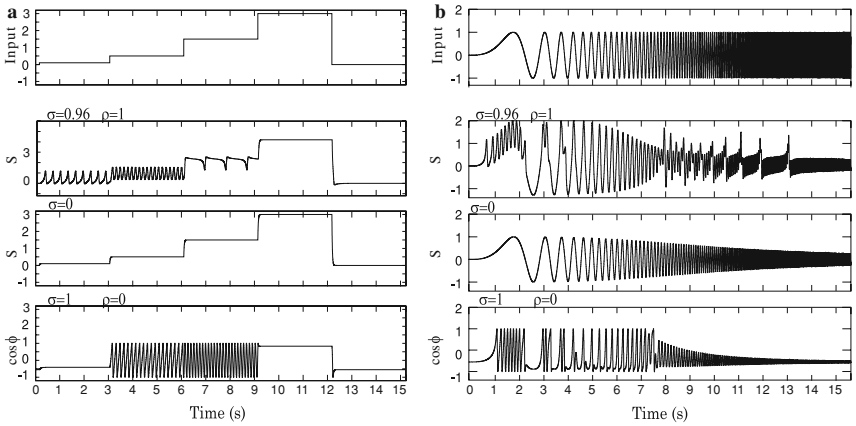


Fig. 4 Voltage response to input currents. **a** for current steps, **b** for oscillatory input of increasing frequency. Upper figures show the input pattern. Then, from up to down, the voltage response is shown for our complex unit, then for a unit without phase modulation, finally, when only the phasic output is considered

Again the two limit cases are considered. First, when $\sigma = 0$ (third row), the unit becomes an integrating unit and the membrane potential oscillates at the input frequency but with exponential attenuation in amplitude. Second, for the phase impact (fourth row), at slow input frequency (lower than ω), one or several oscillations are nested in the up-state. When the input frequency is larger than ω ($t > 650$ ms), oscillations follows the input frequency which leads to faster oscillations with attenuation.

When the two dynamics are coupled (second row), at very low frequencies (lower than 2 Hz), the dynamics goes from fixed point dynamics to limit-cycle dynamics during up-states. We propose that these oscillations at the theta frequency occurring for positive currents could model the increase of activity during 'up-states'. For faster frequencies of the input signal, but still slower than the phase frequency ω, one oscillation occurs during the up-state, leading to higher depolarization of the membrane potential, and an attenuation function of the frequency appears as a result of the integration. At frequencies faster that the phase frequency ω, interference occur and generate higher mode resonances.

Two coupled units

For sake of simplicity, we will discuss results obtained for symmetrical connections ($w_{12} = w_{21} = w$). In the two following simulations, the dynamics of the network is analyzed after a transient excitation is applied to one unit. Figure 5 analyzes the dynamics by looking at the extrema of the membrane potential and by quantifying the phase difference between the two units. The measure of synchrony is obtained by dividing the time between the maximum of the membrane potential of the two units, $\delta\psi = t(S_1^{\max}) - t(S_2^{\max})$, by the period of the signal, T. Accordingly, $\delta\psi/T = 0$ corresponds to the two units oscillating at same phase (phase locked), and $\delta\psi/T = 0.5$ corresponds to the two units being antiphase locked.

The impact of different type of inputs is beyond the scope of this paper and only the impact of a transient input applied to one unit is considered.

From an Hopfield network to a flip flop network

For $\sigma = 0$, the membrane potential is not influenced by the phase and the system of two equations simplifies to a classical recurrent network. Since Hopfield, that system is well known for its ability to encode patterns as associative memories in fixed point attractors (Hopfield 1982; Amit 1989; Durstewitz et al. 2000). In our two units network, if the coupling strength w is strong enough (w larger than 0.68), the dynamics is characterized by two stable fixed points, a down state ($S_1 = S_2 = 0$) and an up-state (visible in Fig. 5a with $S_1 \sim 0.65$). Applying a positive transient input to one of the two cells suffices to move the dynamics from the down-state to the up-state. In return, a negative input can bring back the dynamics to the down state.

When increasing σ, a small oscillation characterizes the up-state and associative memory properties (storage and completion) are preserved. The two cells oscillate at nearly identical phases.

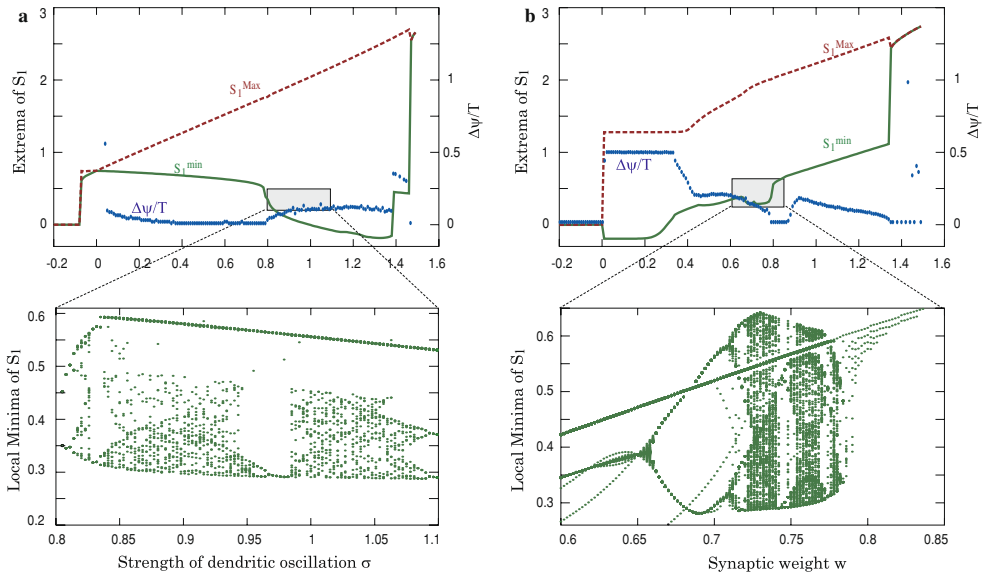


Fig. 5 Extrema of the membrane potential during the up-state ($S_1^{\min}$ and $S_1^{\max}$) and normalized phase difference ($\Delta\psi/T = (\phi_2 - \phi_1)/T$) for a two-unit network when either the internal coupling strength (**a**) or the synaptic weight (**b**) is varied. Lower figures demonstrate the presence of complex dynamics in specific region of the upper figures by showing the bifurcation diagram for the local minima of the membrane potential. For (**a**), $w = 0.75$. For (**b**), $\sigma = 0.9$

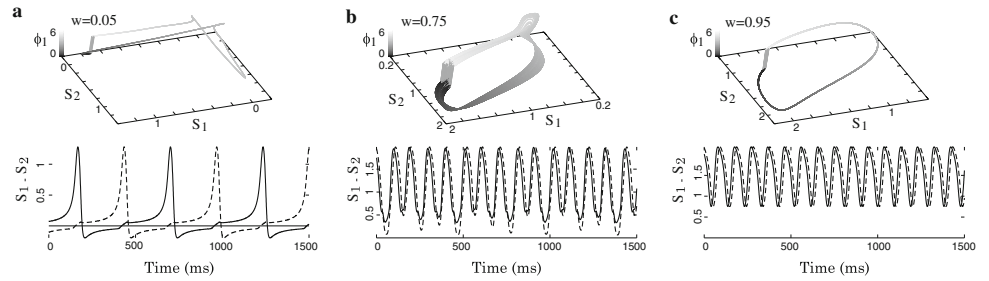


Fig. 6 S_i temporal evolution, (S_1,S_2) phase plane and (S_i, ϕ_i) cylinder space. **a** Up-state oscillation for strong coupling. **b** Multiple frequency oscillation for intermediate coupling. **c** Down-state oscillation for weak coupling

For large values of σ (i.e. large internal coupling), the membrane potential saturates. The region of interest lies before saturation when the membrane returns transiently near the resting state and accordingly not so far from the Milnor attractor. This leads to complex dynamics as shown by the fractal structure of the local minima of S_1 (lower part of Fig. 5a). In that situation, a transient input applied simultaneously to the two units would synchronize the two units' activity and after one oscillation, when the membrane potentials cross the resting state, they stop their activity and remain in the fixed point attractor (not shown here).

Influence of the synaptic weight

Figures 5b and 6 show the impact of varying the synaptic weight on the dynamics of the two units after transient stimulation of one unit.

For small synaptic weights, the two units oscillate anti-phase (Fig. 6a) and their membrane potentials periodically

visit the resting state for a long time. The frequency of this oscillation increases with the synaptic weights and for strong synaptic weights (Fig. 6c), rapid in-phase oscillation appear. In that situation, the oscillation does not cross anymore the fixed point attractor, the dynamics remains in up-state. For intermediate coupling strength, an intermediate cycle is observed and more complex dynamics occur for a small range of weights ($0.58 < w < 0.78$ with our parameters, Fig. 5b). The bifurcation diagram shows multiple roads to chaos through period doubling (Fig. 6b).

Working memory of cell assemblies

In this section, a network of $N = 80$ cells containing eight strongly overlapped cell assemblies is simulated. Each cell assembly contained 10 cells among which seven were overlapping with other cell assemblies. The intersection between two cell assemblies was limited to two cells. The connectivity of the network was chosen bimodal: The synaptic weights between cells lying in the same cell assembly were chosen from the normal distribution ($\mu = 0.8$ and $\sigma = 0.15$), while the other weights were chosen from the normal distribution ($\mu = 0.2$ and $\sigma = 0.1$).

To avoid saturation, a global inhibition unit has been added to the network. This unit computes the total activity of the network ($A = \sum R(S_i)$) and inhibits all cell with the following negative current:

$$\Delta(\gamma(A - \kappa N)) \tag{10}$$

where $\Delta(x) = -x$ for $x > 0$ and 0 elsewhere; γ defines the strength of the inhibitory cell (here 0.1) and κ (in %) defines a threshold triggering the inhibition (here 0.03, meaning that inhibition starts when more than 3% of cells are activated).

Spontaneous activity

Figure 7 shows the spontaneous activity of the network, i.e. in absence of any external stimuli. The upper part of the figure shows the evolution of the membrane potential for the 80 cells (each cell is represented by a different color). The membrane potential is maintained around the Milnor attractor (nearby the stable fixed point). The middle part in Fig. 7 shows rasterplot of cells activity. Global inhibition was tuned to prevent the simultaneous activation of multiple cell assemblies and each period is associated with the activation of a specific subset of cells. Lower part in Fig. 7 quantifies the proportion of each cell assembly activated at each time step. During the observed period, all cell assemblies (characterized in the figure by a different color and letter) are reactivated. It has to be noted that since cell assemblies are overlapping, the total reactivation of a cell assembly is necessarily associated with the partial reactivation of other cell assemblies.

The activity of the network can be explained in the following way: when one cell is activated, it tends to

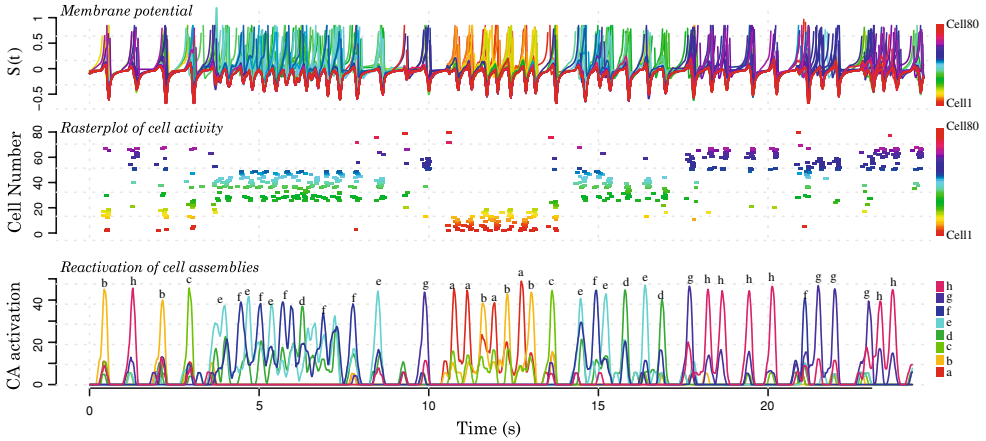


Fig. 7 Spontaneous activity in a 80 units network containing eight overlapped cell assemblies of 10 cells each. The upper and the middle figures show respectively the membrane potential and a rasterplot of the activity of each individual cell (one color per cell). The lower figure shows the reactivation of the different cell assemblies (each assembly has its own color and letter). Periods of no-activity alternate with periods of activity during which specific cell assemblies are preferentially activated

activate the cells belonging to the same cell assembly and the activated cells undergo one oscillation before going back to the resting state, nearby to the Milnor attractor. At the Milnor attractor, any cell can be kicked out of the attractor leading to the activation of a different or the same cell assembly. As a result, the spontaneous activity of the system is characterized by the Milnor attractor which provides a kind of reset of the entire network, enabling the activation of the different stored memories. The system itinerates among the previously stored memories, with the passage from one memory to a new one characterized by the "ruin of the Milnor attractor" leading to a kind of "chaotic itinerancy" (Tsuda 1992; Kaneko 1992).

The overlapping structure of the cell assemblies can influence the sequence of reactivation: The reactivation of one specific cell assembly will tend to reactivate the cell assemblies sharing common cells.

Working memory

In this section, after applying external stimuli during short time periods to part of one or several cell assemblies embedded in the network, we observe the network's ability to sustain the activity of these cell assemblies in agreement with the theory of working memory. To reflect biologically plausible conditions, cell assemblies were sequentially activated during 100 computational time steps; i.e. approximately 200 ms.

In the previous section, we saw that after the reactivation of one cell assembly, all cell assemblies are equally

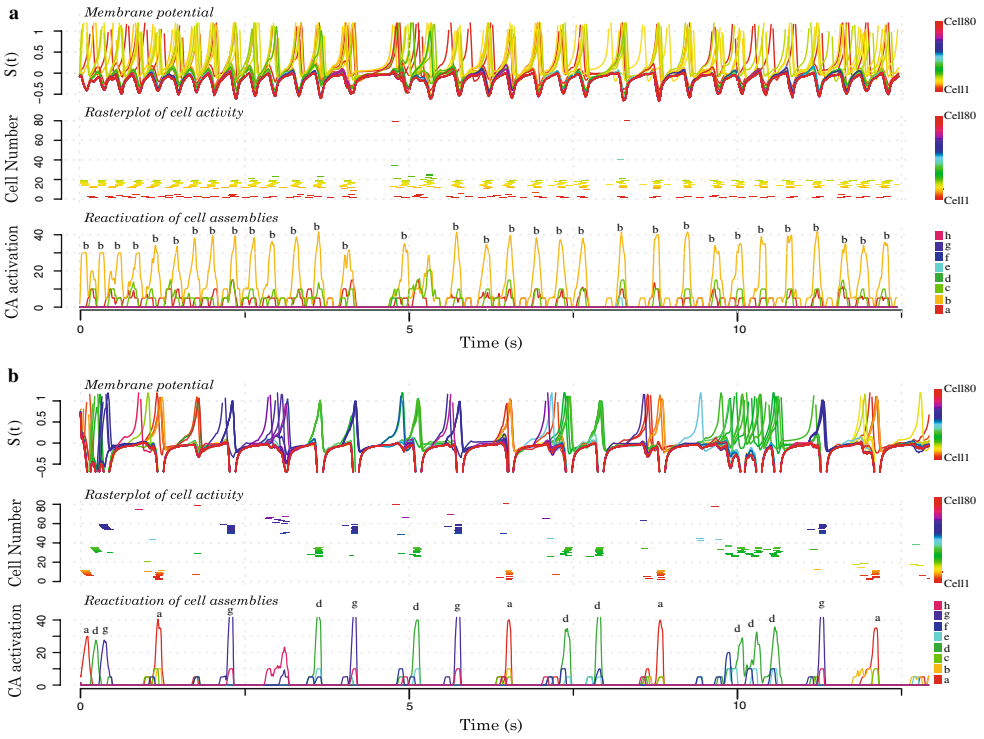


Fig. 8 Working memory in a 80 units network containing eight overlapped cell assemblies of 10 cells each. In both **a** and **b**, upper figures show the membrane potential of each individual cell (each cell one color). Middle figures show rasterplots of individual cells activity (a cell is said to be in an active state if its spike density is larger than 0.5). Lower figures show the reactivation of the different cell assemblies (one color and letter per cell assembly). **a** An external stimulus was impinging part (40%) of one cell assembly during a short transient (10 computational time steps). As a result, this cell assembly is continuously activated as a short term memory. **b** External stimuli are successively applied to part (40%) of three cell assemblies (each CA is stimulated during 100 computational time steps; i.e. approximately 200 ms). After stimulation, we observe that these three cell assemblies have sustained activity

probable to be reactivated. To modify this probability and to increase the probability of selected cell assemblies to be reactivated, as previously proposed (Molter et al. 2008), we simulated a transient period of Hebbian synaptic plasticity during the presentation of the stimuli. Practically, at each computational step, weights between cells reactivated during that period were increased by 0.01. We believe that the proposed mechanism can be related with growing evidence showing that attention-like processes are associated with period of short term plasticity (van Swinderen 2007; Jaaskelainen et al. 2007).

In Fig. 8a, part (40%) of one cell assembly was stimulated. In Fig. 8b, part (40%) of three cell assemblies were sequentially stimulated. First, we observe that in agreement with the theory of associative memories, in both scenario, rasterplot figures indicate that the stimulation of 40% of a cell assembly results in its complete reactivation. Second, it appears that in both cases, the transient stimulation of a cell assembly leads to its preferential reactivation and to a form of sustained activity.

The explanation is the following. After external stimulation, the network is attracted to its resting state, and again the dynamics reflects the presence of diverging orbits crossing attracting orbits. When one cell assembly was reactivated (Fig. 8a), its recurrent connections were increased during the attention-like reactivation, and that cell assembly is now more likely to win the competition and to be reactivated. Since the rapid weight increase due to attention is not balanced by a slow decrease, no dynamics transition is expected and short term memory is expected to last forever (as in Brunel et al. 2004; Mongillo et al. 2008). To implement a slow extinction of the working memory (after dozens of seconds), a mechanism of slow weight decrease could be implemented.

When three cell assemblies were reactivated (Fig. 8b), the balance between excitation and inhibition results in the competition of these three cell assemblies which leads to complex patterns of activity. As a result, the three cell assemblies show clear sustained activity, and these reactivations occur at different times. This result appears important since their simultaneous reactivation would mix them and would prevent any possibility to decode the utile information embedded in the network. Together, these results confirm that our model satisfies one important defining feature of the working memory: The ability to maintain simultaneously multiple chunks of information in short term memory.

Discussion

Recent neuroscience emphasized the importance of the brain dynamics to achieve higher cognitive functions. From a computational perspective, it gave straightforward directions for the development of new type of models where dynamics played more central roles (Kozma and Freeman 2003; Molter et al. 2007). In this paper, to conciliate the classical Hopfield network (Hopfield 1982) with the need of more complex dynamics, we propose a new biologically motivated network model, called “flip-flop oscillation network”.

In this model, each cell is embedded in a recurrent network and is characterized by two internally coupled variables, the cell’s membrane potential and the cell’s phase of firing activity relatively to the theta local field potential.

In a first study, we demonstrate theoretically that in the cylinder space, the Milnor attractor (Milnor 1985) appears at a critical condition through forward and reverse saddle-node bifurcations for a one-cell network (Figs. 1, 2). Near the critical condition, the pair of saddle and node constructs a pseudo-attractor, which leads to Milnor attractor-like properties in computer experiments. Nearby the attractor, the dynamics of the cell is characterized by high sensitivity: infinitesimal transient perturbation can activate an oscillation of the membrane potential.

Simulations of a two-cell network revealed the presence of numerous complex dynamics. We observed that semi-stability of the Milnor attractor dynamics characterizing one cell dynamics, combined in a two cell-network leads to oscillations and chaotic dynamics through period doubling roads (Fig. 5). The important role played by the Milnor attractor for the apparition of the chaotic attractor suggests chaotic itinerancy (Tsuda 2001).

Finally, we tested our model during spontaneous activity (Fig. 7) and for selective maintenance of previously stored information (Fig. 8). In agreement with the cell assembly theory (Hebb 1949), multiple overlapping cell assemblies were phenomenologically embedded in the network to simulate chunks of information.

During spontaneous activity, the network is dynamically switching between the different cell assemblies. After the reactivation of a specific cell assembly, the quasi-stable state of the Milnor attractor provides receptivity to second order dynamics of the internal state and to external stimuli. During that “receptive” or “attention-like” state, a different information can be reactivated. This is supported by recent biological reports; for example, the spatiotemporal pattern of activity of cortical neurons observed during thalamically triggered events are similar to the ones observed during spontaneous events (MacLean et al. 2005). From a dynamics perspective, the presence of the Milnor attractor prevents the network to be “occulted” by one information and leads to chaotic itinerancy between the previously stored cell assemblies.

To simulate working memory tasks, we analyzed if the network could transiently hold specific memories triggered

by external stimuli. To reliably and robustly enforce the maintenance of triggered cell assemblies, short period of synaptic plasticity was simulated, reproducing attention-like processes (van Swinderen 2007; Jaaskelainen et al. 2007). As a result, one or multiple cell assemblies could selectively be maintained. An important feature brought by the addition of the theta rhythm to the classical Hopfield network is that different cell assemblies are separately reactivated at different theta oscillations. At each theta cycle, cells from a unique cell assembly are synchronously reactivated. That enables our network to solve the binding problem by reactivating at different phases different overlapping cell assemblies.

If the addition of an internal oscillation to classical rate coding model (e.g. Brunel et al. 2004; Mongillo et al. 2008) can solve the binding problem, it can still not explain in an easy way the limited memory capacity of working memory [e.g. the magical number seven (Miller 1956)]. In that sense, it differs from the seminal paper from Idiart et al. (Lisman and Idiart 1995) where it was proposed that memories are stored in gamma cycles embedded in theta cycles, and that the magical number seven is explained by the number of gamma cycles which can be embedded in a theta cycle. Further studies should focus more deeply on the capacity problem. More precisely, the reactivation of multiple items at different phase of a same theta cycle should be tested.

To summarize, we are proposing a compact, effective, and powerful "flip-flop oscillations network" whose dynamical complexity can be of interest for further analysis of integrative brain dynamics. First attempts to solve working memory tasks gave promising results were different chunks of information were reactivated at different theta cycles. Finally, in our model, the information conveyed during spontaneous activity and working memory tasks appeared similar, reminding the famous quote from Rodolfo Llinas: "A person's waking life is a dream modulated by the senses" (Llineas 2001).

References

Amari S (1977) Neural theory of association and concept-formation. Biol Cybern 26:175–185

Amit DJ (1989) Modelling brain function: the world of attractor networks. Cambridge University Press, New York

Babloyantz A, Destexhe A (1986) Low-dimensional chaos in an instance of epilepsy. Proc Natl Acad Sci USA 83:3513–3517

Baddeley AD (1986) Working memory. Oxford University Press, New York

Brunel N, Hakim V, Isope P, Nadal JP, Barbour B (2004) Optimal information storage and the distribution of synaptic weights; perceptron versus purkinje cell. Neuron 43:745–757

Buzsaki G, Draguhn A (2004) Neuronal oscillations in cortical networks. Science 304:1926–1929

Colliaux D, Yamaguchi Y, Molter C, Wagatsuma H (2007) Working memory dynamics in a flip-flop oscillations network model with milnor attractor. In: Proceedings of ICONIP, Kyoto, Japan

Durstewitz D, Seamans JK, Sejnowski TJ (2000) Neurocomputational models of working memory. Nat Neurosci 3:1184–1191

Fuster JM, Alexander GE (1971) Neuron activity related to short-term memory. Science 173(997):652–654

Gianluigi M, Barak O, Tsodyks M (2008) Synaptic theory of working memory. Science 319(5869):1543–1546

Goldman-Rakic PS (1995) Models of information processing in the basal ganglia. MIT Press, Cambridge, pp 131–148

Grossberg S (1992) Neural networks and natural intelligence. MIT Press, Cambridge

Hebb DO (1949) The organization of behavior; a neuropsychological theory. Wiley, New York

Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. Proc Natl Acad Sci USA 79:2554–2558

Jaaskelainen IP, Ahveninen J, Belliveau JW, Raij T, Sams M (2007) Short-term plasticity in auditory cognition. Trends Neurosci 30(12):653–661

Kaneko K (1992) Pattern dynamics in spatiotemporal chaos. Physica D 34:141

Kaneko K (2002) Dominance of milnor attractors in globally coupled dynamical systems with more than 7 ± 2 degrees of freedom. Phys Rev E 66:055201

Kenet T, Bibitchkov D, Tsodyks M, Grinvald A, Arieli A (2003) Spontaneously emerging cortical representations of visual attributes. Nature 425:954–956

Kozma R, Freeman WJ (2003) Basic principles of the kiv model and its application to the navigation problem. J Integr Neurosci 2(1):125–145

Lisman JE, Idiart MA (1995) Storage of 7 ± 2 short-term memories in oscillatory subcycles. Science 267:1512–1516

Llineas R (2001) I of the vortex: from neurons to self. MIT Press, Cambridge

MacLean JN, Watson BO, Aaron GB, Yuste R (2005) Internal dynamics determine the cortical response to thalamic stimulation. Neuron 48:811–823

McCormick DA (2005) Neuronal networks: flip-flops in the brain. Curr Biol 15(8):294–296

Milnor J (1985) On the concept of attractor. Commun Math Phys 99(102):177–195

Mizuhara H, Yamaguchi Y (2007) Human cortical circuits for central executive function emerge by theta phase synchronization. Neuroimage 36(361):232–244

Molter C, Salihoglu U, Bersini H (2007a) The road to chaos by time asymmetric hebbian learning in recurrent neural networks. Neural Comput 19(1):100

Molter C, Salihoglu U, and Bersini H (2007b) Neurodynamics of higher level cognition and consciousness. In: Giving meaning to cycles to go beyond the limitations of fixed point attractors. Springer Verlag

Molter C, Sato N, Yamaguchi Y (2007c) Reactivation of behavioral activity during sharp waves: a computational model for two stage hippocampal dynamics. Hippocampus 17(3):201–209

Molter C, Colliaux D, Yamaguchi Y (2008) Working memory and spontaneous activity of cell assemblies. A biologically motivated computational model. In: Proceedings of the IJCNN conference. IEEE Press, Hong Kong, pp 3069–3076

Miller GA (1956) The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychol Rev 63:81–97

Muresan RC, Savin C (2007) Resonance or integration? Self-sustained dynamics and excitability of neural microcircuits. J Neurophysiol 97:1911–1930

Nicolis J, Tsuda I (1985) Chaotic dynamics of information processing: the "magic number seven plus-minus two" revisited. Bull Math Biol 47:343–365

Onton J, Delorme A, Makeig S (2005) Frontal midline eeg dynamics during working memory. Neuroimage 27(2):341–356

Ott E (1993) Chaos in dynamical systems. Cambridge University-Press, Cambridge.

Rainer G, Asaad WF, Miller EK (1998) Selective representation of relevant information by neurons in the primate prefrontal cortex. Nature 393:577–579

Rainer G, Lee H, Simpson GV, Logothetis NK (2004) Working-memory related theta (4–7 Hz) frequency oscillations observed in monkey extrastriate visual cortex. Neurocomputing 58–60: 965–969

Skarda CA, Freeman W (1987) How brains make chaos in order to make sense of the world. Behav Brain Sci 10:161–195

Tsuda I (1992) Dynamic link of memorychaotic memory map in nonequilibrium neural networks. Neural Netw 5:313–326

Tsuda I (2001) Towards an interpretation of dynamic neural activity in terms of chaotic dynamical systems. Behav Brain Sci 24:793–810

Tsujimoto T, Shimazu H, Isomura Y, Sasaki K (2003) Prefrontal theta oscillations associated with hand movements triggered by warning and imperative stimuli in the monkey. Neurosci Lett 351(2): 103–106

van Swinderen B (2007) Attention-like processes in drosophila require short-term memory genes. Science 315(5818):1590–1593

Varela F, Lachaux J-P, Rodriguez E, Martinerie J (2001) The brainweb: phase synchronization and large-scale integration. Nat Rev Neurosci 2:229–239

Whitlock JR, Heynen AJ, Shuler MG, Bear MF (2006) Learning induces long-term potentiation in the hippocampus. Science 313(5790):1093–1097

Yamaguchi Y (2003) A theory of hippocampal memory based on theta phase precession. Biol Cybern 89:1–9

Bibliography

[1] C. C. Hilgetag, M. A. O'Neill, and M. P. Young. Hierarchical organization of macaque and cat cortical sensory systems explored with a novel network processor. *Philos Trans R Soc Lond B Biol Sci*, 355(1393):71–89, Jan 2000.

[2] Ed Bullmore and Olaf Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat Rev Neurosci*, 10(3):186–198, Mar 2009.

[3] Stephen D Van Hooser. Similarity and diversity in visual cortex: is there a unifying theory of cortical computation? *Neuroscientist*, 13(6):639–656, Dec 2007.

[4] Manabu Tanifuji, Kazushige Tsunoda, and Yukako Yamane. Representation of object images by combinations of visual features in the macaque inferior temporal cortex. *Novartis Found Symp*, 270:217–25; discussion 226–37, 2006.

[5] Pasko Rakic. Confusing cortical columns. *Proc Natl Acad Sci U S A*, 105(34):12099–12100, Aug 2008.

[6] P. Reinagel and R. C. Reid. Temporal coding of visual information in the thalamus. *J Neurosci*, 20(14):5392–5400, Jul 2000.

[7] Rufin VanRullen, Rudy Guyonneau, and Simon J Thorpe. Spike times make sense. *Trends Neurosci*, 28(1):1–4, Jan 2005.

[8] C. M. Gray, P. Knig, A. K. Engel, and W. Singer. Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature*, 338(6213):334–337, Mar 1989.

[9] Gyorgy Buzsáki and Andreas Draguhn. Neuronal oscillations in cortical networks. *Science*, 304(5679):1926–1929, Jun 2004.

[10] P. S. Goldman-Rakic. Cellular basis of working memory. *Neuron*, 14(3):477–485, Mar 1995.

[11] David A McCormick. Neuronal networks: flip-flops in the brain. *Curr Biol*, 15(8):R294–R296, Apr 2005.

[12] D. Durstewitz and J. K. Seamans. Beyond bistability: biophysics and temporal dynamics of working memory. *Neuroscience*, 139(1):119–133, Apr 2006.

[13] M. Steriade, I. Timofeev, and F. Grenier. Natural waking and sleep states: a view from inside neocortical neurons. *J Neurophysiol*, 85(5):1969–1985, May 2001.

[14] Henri Korn and Philippe Faure. Is there chaos in the brain? ii. experimental evidence and related models. *C R Biol*, 326(9):787–840, Sep 2003.

[15] Mikhail I. Rabinovich, Pablo Varona, Allen I. Selverston, and Henry D. I. Abarbanel. Dynamical principles in neuroscience. *Reviews of Modern Physics*, 78(4):1213–1265, November 2006.

[16] Ichiro Tsuda. Toward an interpretation of dynamic neural activity in terms of chaotic dynamical sys- tems. *Behavioral and Brain Sciences*, 24(5):793, 2001.

[17] Misha Rabinovich, Ramon Huerta, and Gilles Laurent. Neuroscience. transient dynamics for neural processing. *Science*, 321(5885):48–50, Jul 2008.

[18] F. Varela, J. P. Lachaux, E. Rodriguez, and J. Martinerie. The brainweb: phase synchronization and large-scale integration. *Nat Rev Neurosci*, 2(4):229–239, Apr 2001.

[19] Andrea Benucci, Robert A Frazor, and Matteo Carandini. Standing waves and traveling waves distinguish two circuits in visual cortex. *Neuron*, 55(1):103–117, Jul 2007.

[20] Walter J Freeman. Vortices in brain activity: their mechanism and significance for perception. *Neural Netw*, 22(5-6):491–501, 2009.

[21] Marcus E Raichle. Two views of brain function. *Trends Cogn Sci*, 14(4):180–190, Apr 2010.

[22] Richard Gregory. *The intelligent eye.* 1970.

[23] Michael D Fox and Marcus E Raichle. Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nat Rev Neurosci*, 8(9):700–711, Sep 2007.

[24] Alain Destexhe, Michael Rudolph, and Denis Par. The high-conductance state of neocortical neurons in vivo. *Nat Rev Neurosci*, 4(9):739–751, Sep 2003.

[25] Tal Kenet, Dmitri Bibitchkov, Misha Tsodyks, Amiram Grinvald, and Amos Arieli. Spontaneously emerging cortical representations of visual attributes. *Nature*, 425(6961):954–956, Oct 2003.

[26] Eugen M. Izhikevich. *Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting.* MIT Press, 2007.

[27] Stefan Haeusler and Wolfgang Maass. A statistical analysis of information-processing properties of lamina-specific cortical microcircuit models. *Cereb Cortex*, 17(1):149–162, Jan 2007.

[28] A. Lansner and M. Lundqvist. *Dynamic Coordination in the Brain: From Neurons to Mind. Strngmann Forum Report,volume 5*, chapter Modeling coordination in the neocortex at the microcircuit and global network level, pages pages 83–99. MIT Press, 2010.

[29] J. J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proc Natl Acad Sci U S A*, 79(8):2554–2558, Apr 1982.

[30] M. Diesmann, M. O. Gewaltig, and A. Aertsen. Stable propagation of synchronous spiking in cortical neural networks. *Nature*, 402(6761):529–533, Dec 1999.

[31] Eugene M Izhikevich. Polychronization: computation with spikes. *Neural Comput*, 18(2):245–282, Feb 2006.

[32] Wolfgang Maass, Thomas Natschlger, and Henry Markram. Real-time computing without stable states: a new framework for neural computation based on perturbations. *Neural Comput*, 14(11):2531–2560, Nov 2002.

[33] C. van Vreeswijk and H. Sompolinsky. Chaotic balanced state in a model of cortical circuits. *Neural Comput*, 10(6):1321–1371, Aug 1998.

[34] Arvind Kumar, Sven Schrader, Ad Aertsen, and Stefan Rotter. The high-conductance state of cortical networks. *Neural Comput*, 20(1):1–43, Jan 2008.

[35] Alfonso Renart, Jaime de la Rocha, Peter Bartho, Liad Hollender, Nstor Parga, Alex Reyes, and Kenneth D Harris. The asynchronous state in cortical circuits. *Science*, 327(5965):587–590, Jan 2010.

[36] Néstor Parga and Larry F Abbott. Network model of spontaneous activity exhibiting synchronous transitions between up and down states. *Front Neurosci*, 1(1):57–66, Nov 2007.

[37] Alain Destexhe. Self-sustained asynchronous irregular states and up-down states in thalamic, cortical and thalamocortical networks of nonlinear integrate-and-fire neurons. *J Comput Neurosci*, 27(3):493–506, Dec 2009.

[38] Bruno Cessac, Héléne Paugam-Moisy, and Thierry Viville. Overview of facts and issues about neural coding by spikes. *J Physiol Paris*, 104(1-2):5–18, 2010.

[39] M. Carandini and D. Ferster. Membrane potential and firing rate in cat primary visual cortex. *J Neurosci*, 20(1):470–484, Jan 2000.

[40] F. Rieke, D .Warland, R. de Ruyter van Steveninck, and W. Bialek. *Spikes: Exploring the Neural Code.* MIT Press, 1997.

[41] F. Gabbiani and C. Koch. *Methods in Neuronal Modeling: From Ions to Networks.*, chapter Principles of spike train analysis., pages pp. 313–360. MIT Press, 1998.

[42] W. Bair and C. Koch. Temporal precision of spike trains in extrastriate cortex of the behaving macaque monkey. *Neural Comput*, 8(6):1185–1202, Aug 1996.

[43] M. Nawrot, A. Aertsen, and S. Rotter. Single-trial estimation of neuronal firing rates: from single-neuron spike trains to population activity. *J Neurosci Methods*, 94(1):81–92, Dec 1999.

[44] Yonatan Loewenstein, Sverine Mahon, Paul Chadderton, Kazuo Kitamura, Haim Sompolinsky, Yosef Yarom, and Michael Husser. Bistability of cerebellar purkinje cells modulated by sensory stimulation. *Nat Neurosci*, 8(2):202–211, Feb 2005.

[45] J. A. Hartigan and P. M. Hartigan. The dip test of unimodality. *Annals of Statistics*, 13:70–84, 1985.

[46] James W. Cooley and John W. Tukey. An algorithm for the machine calculation of complex fourier series. *Math. Comput.*, 19:297301, 1965.

[47] Gerhard Werner. Fractals in the nervous system: conceptual implications for theoretical neuroscience. *Frontiers in Physiology*, (0):12, 2010.

[48] Sami El Boustani, Olivier Marre, Sbastien Béhuret, Pierre Baudot, Pierre Yger, Thierry Bal, Alain Destexhe, and Yves Frégnac. Network-state modulation of power-law frequency-scaling in visual cortical neurons. *PLoS Comput Biol*, 5(9):e1000519, Sep 2009.

[49] Christopher Torrence and Gilbert P. Compo. A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society*, Vol. 79, 1:pp. 6178, 1998.

[50] Stéphane Mallat. *A Wavelet Tour of Signal Processing, 2nd Edition.* Academic Press, 1999.

[51] Franois B Vialatte, Jordi Sol-Casals, Justin Dauwels, Monique Maurice, and Andrzej Cichocki. Bump time-frequency toolbox: a toolbox for time-frequency oscillatory bursts extraction in electrophysiological signals. *BMC Neurosci*, 10:46, 2009.

[52] Claire Sergent, Sylvain Baillet, and Stanislas Dehaene. Timing of the brain events underlying access to consciousness during the attentional blink. *Nat Neurosci*, 8(10):1391–1400, Oct 2005.

[53] Sam M Doesburg, Alexa B Roggeveen, Keiichi Kitajo, and Lawrence M Ward. Large-scale gamma-band phase synchronization and selective attention. *Cereb Cortex*, 18(2):386–396, Feb 2008.

[54] Damien Gervasoni, Shih-Chieh Lin, Sidarta Ribeiro, Ernesto S Soares, Janaina Pantoja, and Miguel A L Nicolelis. Global forebrain dynamics predict rat behavioral states and their transitions. *J Neurosci*, 24(49):11137–11147, Dec 2004.

[55] J.D. Victor and K. Purpura. Metric-space analysis of spike trains: theory, algorithms, and application. *Network*, 8:127–164, 1997.

[56] M. C. van Rossum. A novel spike distance. *Neural Comput*, 13(4):751–763, Apr 2001.

[57] S. Schreiber, J. M. Fellous, D. Whitmer, P. Tiesinga, and T. J. Sejnowski. A new correlation-based measure of spike timing reliability. *Neurocomputing*, 52-54:925–931, Jun 2003.

[58] Thomas Kreuz, Julie S Haas, Alice Morelli, Henry D I Abarbanel, and Antonio Politi. Measuring spike train synchrony. *J Neurosci Methods*, 165(1):151–161, Sep 2007.

[59] Joy A. Thomas Thomas M. Cover. *Elements of Information Theory, 2nd Edition.* Wiley, 2006.

[60] Shunichi Amari and Hiroshi Nagaoka. Methods of information geometry. *Translations of mathematical monographs, American Mathematical Society*, 191, 2000.

[61] G. Tononi, O. Sporns, and G. M. Edelman. A measure for brain complexity: relating functional segregation and integration in the nervous system. *Proc Natl Acad Sci U S A*, 91(11):5033–5037, May 1994.

[62] L. Barnett, C. L. Buckley, and S. Bullock. Neural complexity and structural connectivity. *Phys Rev E Stat Nonlin Soft Matter Phys*, 79(5 Pt 1):051914, May 2009.

[63] N. Bertschinger N. Ay, E. Olbrich and J. Jost. A unifying framework for complexity measures of finite systems,. In *Proceedings ECCS*, 2006.

[64] Ilya Nemenman, William Bialek, and Rob de Ruyter van Steveninck. Entropy and information in neural spike trains: progress on the sampling problem. *Phys Rev E Stat Nonlin Soft Matter Phys*, 69(5 Pt 2):056111, May 2004.

[65] Robin A. Ince, Alberto Mazzoni, Rasmus S Petersen, and Stefano Panzeri. Open source tools for the information theoretic analysis of neural data. *Frontiers in Neuroscience*, (0):5, 2010.

[66] James A. Yorke Tim Sauer and Martin Casdagli. Embedology. *Journal of Statistical Physics*, 65, Numbers 3-4:579–616, 1991.

[67] Peter Grassberger and Itamar Procaccia. Measuring the strangeness of strange attractors. *Physica D: Nonlinear Phenomena*, Volume 9, Issues 1-2:189–208, 1983.

[68] Rainer Hegger, Holger Kantz, and Thomas Schreiber. Practical implementation of nonlinear time series methods: The tisean package. *Chaos*, 9(2):413–435, Jun 1999.

[69] A. Babloyantz and A. Destexhe. Low-dimensional chaos in an instance of epilepsy. *Proc Natl Acad Sci U S A*, 83(10):3513–3517, May 1986.

[70] David G. Stork Richard O. Duda, Peter E. Hart. *Pattern Classification, 2nd edition.* Wiley-Interscience, 2001.

[71] David Arthur and Sergei Vassilvitskii. k-means++: The advantages of careful seeding. Technical Report 2006-13, Stanford InfoLab, June 2006.

[72] Peter Langfelder, Bin Zhang, and Steve Horvath. Defining clusters from a hierarchical cluster tree: the dynamic tree cut package for r. *Bioinformatics*, 24(5):719–720, Mar 2008.

[73] Teuvo Kohonen. Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43(1):59–69, 1982.

[74] Alain Destexhe, Stuart W Hughes, Michelle Rudolph, and Vincenzo Crunelli. Are corticothalamic 'up' states fragments of wakefulness? *Trends Neurosci*, 30(7):334–342, Jul 2007.

[75] Bilal Haider, Alvaro Duque, Andrea R Hasenstaub, Yuguo Yu, and David A McCormick. Enhancement of visual responsiveness by spontaneous local network activity in vivo. *J Neurophysiol*, 97(6):4186–4202, Jun 2007.

[76] Y. Frégnac, M. Blatow, J.-P. Changeux, J. de Felipe, W. Maass A. Lansner, D. A. McCormick, C. M. Michel, H. Monyer, E. Szathmary, and R. Yuste. Ups and downs in cortical computation. *The Interface between Neurons and Global Brain Function, Dahlem Workshop Report 93*, pages pages 393–433, 2006.

[77] W. McCulloch and W. Pitts. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 7:115 – 133, 1943.

[78] Daniel P Buxhoeveden and Manuel F Casanova. The minicolumn hypothesis in neuroscience. *Brain*, 125(Pt 5):935–951, May 2002.

[79] Jennifer S Lund, Alessandra Angelucci, and Paul C Bressloff. Anatomical substrates for functional columns in macaque monkey primary visual cortex. *Cereb Cortex*, 13(1):15–24, Jan 2003.

[80] Henry Markram. The blue brain project. *Nat Rev Neurosci*, 7(2):153–160, Feb 2006.

[81] V. B. Mountcastle. The columnar organization of the neocortex. *Brain*, 120 (Pt 4):701–722, Apr 1997.

[82] D. H. Hubel and T. N. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J Physiol*, 160:106–154, Jan 1962.

[83] Yuri Kuznetsov. *Elements of applied bifurcation theory.* 2004.

[84] Stephen Wiggins. *Introduction to Applied Nonlinear Dynamical Systems and Chaos.* 1996.

[85] Jurgen Jost. *Dynamical Systems: Examples of Complex Behaviour.* Introduction to Applied Nonlinear Dynamical Systems and Chaos., 2005.

[86] Crispin Gardiner. *Handbook of Stochastic Methods: for Physics, Chemistry and the Natural Sciences.* 1985.

[87] Hannes Risken. *The FokkerPlanck Equation: Methods of Solutions and Applications.* 1989.

[88] Ludwig Arnold. *Random Dynamical Systems.* Springer, 1998.

[89] Giancarlo Benettin, Luigi Galgani, Antonio Giorgilli, and Jean-Marie Strelcyn. Lyapunov characteristic exponents for smooth dynamical systems and for hamiltonian systems; a method for computing all of them. part 1: Theory. *Meccanica*, 15:9–20, 1980. 10.1007/BF02128236.

[90] Petilla Interneuron Nomenclature Group, Giorgio A Ascoli, Lidia Alonso-Nanclares, Stewart A Anderson, German Barrionuevo, Ruth Benavides-Piccione, Andreas Burkhalter, Gyrgy Buzski, Bruno Cauli, Javier Defelipe, Alfonso Fairn, Dirk Feldmeyer, Gord Fishell, Yves Fregnac, Tamas F Freund, Daniel Gardner, Esther P Gardner, Jesse H Goldberg, Moritz Helmstaedter, Shaul Hestrin, Fuyuki Karube, Zoltn F Kisvrday, Bertrand Lambolez, David A Lewis, Oscar Marin, Henry Markram, Alberto Muoz, Adam Packer, Carl C H Petersen, Kathleen S Rockland, Jean Rossier, Bernardo Rudy, Peter Somogyi, Jochen F Staiger, Gabor Tamas, Alex M Thomson, Maria Toledo-Rodriguez, Yun Wang, David C West, and Rafael Yuste. Petilla terminology: nomenclature of features of gabaergic interneurons of the cerebral cortex. *Nat Rev Neurosci*, 9(7):557–568, Jul 2008.

[91] B. W. Connors and M. J. Gutnick. Intrinsic firing patterns of diverse neocortical neurons. *Trends Neurosci*, 13(3):99–104, Mar 1990.

[92] L. F. Abbott, J. A. Varela, K. Sen, and S. B. Nelson. Synaptic depression and cortical gain control. *Science*, 275(5297):220–224, Jan 1997.

[93] G. Q. Bi and M. M. Poo. Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *J Neurosci*, 18(24):10464–10472, Dec 1998.

[94] M. Sterpu and A. Georgescu. Codimension three bifurcations for the fitzhugh- nagumo system. *Mathematical Reports, Acad. Rom.*, 3 (53), 3:p. 287–292, 2001.

[95] Kazuyuki Aihara and Hideyuki Suzuki. Theory of hybrid dynamical systems and its applications to biological and medical systems. *Philos Transact A Math Phys Eng Sci*, 368(1930):4893–4914, Nov 2010.

[96] R. B. Stein. Some models of neuronal variability. *Biophys J*, 7(1):37–68, Jan 1967.

[97] P. Lánský. On approximations of stein's neuronal model. *J Theor Biol*, 107(4):631–647, Apr 1984.

[98] Nicolas Fourcaud and Nicolas Brunel. Dynamics of the firing probability of noisy integrate-and-fire neurons. *Neural Comput*, 14(9):2057–2110, Sep 2002.

[99] Eugene M Izhikevich. Which model to use for cortical spiking neurons? *IEEE Trans Neural Netw*, 15(5):1063–1070, Sep 2004.

[100] Romain Brette and Wulfram Gerstner. Adaptive exponential integrate-and-fire model as an effective description of neuronal activity. *J Neurophysiol*, 94(5):3637–3642, Nov 2005.

[101] Jonathan Touboul and Romain Brette. Dynamics and bifurcations of the adaptive exponential integrate-and-fire model. *Biol Cybern*, 99(4-5):319–334, Nov 2008.

[102] Martin Pospischil, Maria Toledo-Rodriguez, Cyril Monier, Zuzanna Piwkowska, Thierry Bal, Yves Frégnac, Henry Markram, and Alain Destexhe. Minimal hodgkin-huxley type models for different classes of cortical and thalamic neurons. *Biol Cybern*, 99(4-5):427–441, Nov 2008.

[103] Cyrille Rossant, Dan F M Goodman, Jonathan Platkiewicz, and Romain Brette. Automatic fitting of spiking neuron models to electrophysiological recordings. *Front Neuroinformatics*, 4:2, 2010.

[104] K. Tsunoda, Y. Yamane, M. Nishizaki, and M. Tanifuji. Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns. *Nat Neurosci*, 4(8):832–838, Aug 2001.

[105] B. H. Jansen and V. G. Rit. Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. *Biol Cybern*, 73(4):357–366, Sep 1995.

[106] Stephen Grossberg. Towards a unified theory of neocortex: laminar cortical circuits for vision and cognition. *Prog Brain Res*, 165:79–104, 2007.

[107] Alex M Thomson, David C West, Yun Wang, and A. Peter Bannister. Synaptic connections and small circuits involving excitatory and inhibitory neurons in layers 2-5 of adult rat and cat neocortex: triple intracellular recordings and biocytin labelling in vitro. *Cereb Cortex*, 12(9):936–953, Sep 2002.

[108] Tom Binzegger, Rodney J Douglas, and Kevan A C Martin. A quantitative map of the circuit of cat primary visual cortex. *J Neurosci*, 24(39):8441–8453, Sep 2004.

[109] Alex M Thomson and Christophe Lamy. Functional maps of neocortical local circuitry. *Frontiers in Neuroscience*, 5:12, 2007.

[110] Gustavo Deco, Mar Pérez-Sanagustín, Victor de Lafuente, and Ranulfo Romo. Perceptual detection as a dynamical bistability phenomenon: a neurocomputational correlate of sensation. *Proc Natl Acad Sci U S A*, 104(50):20073–20077, Dec 2007.

[111] Hugh R. Wilson and Jack D. Cowan. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical Journal*, Volume 12, Issue 1:1–24, 1972.

[112] N. Brunel. Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *J Comput Neurosci*, 8(3):183–208, 2000.

[113] Tim P Vogels and L. F. Abbott. Signal propagation and logic gating in networks of integrate-and-fire neurons. *J Neurosci*, 25(46):10786–10795, Nov 2005.

[114] Oleksandr V. Popovych, Yuri L. Maistrenko, and Peter A. Tass. Phase chaos in coupled oscillators. *Phys. Rev. E*, 71(6):065201, Jun 2005.

[115] Juan A. Acebrón, L. L. Bonilla, Conrad J. Pérez Vicente, Félix Ritort, and Renato Spigler. The kuramoto model: A simple paradigm for synchronization phenomena. *Rev. Mod. Phys.*, 77(1):137–185, Apr 2005.

[116] David Colliaux, Colin Molter, and Yoko Yamaguchi. Working memory dynamics and spontaneous activity in a flip-flop oscillations network model with a milnor attractor. *Cogn Neurodyn*, 3(2):141–151, Jun 2009.

[117] Douglas Zhou, Yi Sun, Aaditya V Rangan, and David Cai. Spectrum of lyapunov exponents of non-smooth dynamical systems of integrate-and-fire type. *J Comput Neurosci*, 28(2):229–245, Apr 2010.

[118] S. Olmi, A. Politi, and A. Torcini. Collective chaos in pulse-coupled neural networks. *EPL (Europhysics Letters)*, 92(6):60007, 2010.

[119] H. Paugam-Moisy and S.M. Bohte. *Handbook of Natural Computing*, chapter Computing with Spiking Neuron Networks. 2009.

[120] Paul C Bressloff and Jack D Cowan. A spherical model for orientation and spatial-frequency tuning in a cortical hypercolumn. *Philos Trans R Soc Lond B Biol Sci*, 358(1438):1643–1667, Oct 2003.

[121] Geoffrey J Goodhill. Contributions of theoretical modeling to the understanding of neural map development. *Neuron*, 56(2):301–311, Oct 2007.

[122] Richard Durbin, Richard Szeliski, and Alan Yuille. An analysis of the elastic net approach to the traveling salesman problem. *Neural Comput.*, 1:348–358, September 1989.

[123] R. Durbin and D. Willshaw. An analogue approach to the travelling salesman problem using an elastic net method. *Nature*, 326(6114):689–691, 1987.

[124] D.J. Goodhill, G.J. & Willshaw. Application of the elastic net algorithm to the formation of ocular dominance stripes. *Network*, 1:41–59, 1990.

[125] Geoffrey J. Goodhill and David J. Willshaw. Elastic net model of ocular dominance: Overall stripe pattern and monocular deprivation. *Neural Computation*, 6:615–621, 1994.

[126] M. C. Cross and P. C. Hohenberg. Pattern formation outside of equilibrium. *Rev. Mod. Phys.*, 65(3):851, Jul 1993.

[127] J.D. Murray. *Mathematical Biology: II. Spatial Models and Biomedical Applications*. Springer-Verlag, 3rd ed., 2002.

[128] N. V. Swindale. A model for the formation of ocular dominance stripes. *Proc R Soc Lond B Biol Sci*, 208(1171):243–264, Jun 1980.

[129] Paul C Bressloff and Andrew M Oster. Theory for the alignment of cortical feature maps during development. *Phys Rev E Stat Nonlin Soft Matter Phys*, 82(2 Pt 1):021920, Aug 2010.

[130] N. V. Swindale. A model for the formation of orientation columns. *Proc R Soc Lond B Biol Sci*, 215(1199):211–230, May 1982.

[131] F. Wolf and T. Geisel. Spontaneous pinwheel annihilation during visual development. *Nature*, 395(6697):73–78, Sep 1998.

[132] Matthias Kaschube, Michael Schnabel, and Fred Wolf. Self-organization and the selection of pinwheel density in visual cortical development. *New Journal of Physics*, 10(1):015009, 2008.

[133] Ha Youn Lee, Mehdi Yahyanejad, and Mehran Kardar. Symmetry considerations and development of pinwheels in visual maps. *Proceedings of the National Academy of Sciences of the United States of America*, 100(26):16036–16040, 2003.

[134] P. C. Bressloff, J. D. Cowan, M. Golubitsky, P. J. Thomas, and M. C. Wiener. Geometric visual hallucinations, euclidean symmetry and the functional architecture of striate cortex. *Philos Trans R Soc Lond B Biol Sci*, 356(1407):299–330, Mar 2001.

[135] M. Golubitsky and I. Stewart. *The Symmetry Perspective: From Equilibrium to Chaos in Phase Space and Physical Space.* Birkhauser, 2002.

[136] Ian H Stevenson, Beau Cronin, Mriganka Sur, and Konrad P Kording. Sensory adaptation and short term plasticity as bayesian correction for a changing brain. *PLoS One*, 5(8):e12436, 2010.

[137] T. V. Bliss and T. Lomo. Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *J Physiol*, 232(2):331–356, Jul 1973.

[138] D.O. Hebb. *The Organization of Behavior: A Neuropsychological Theory.* 1949.

[139] Geoffrey E. Hinton & Ronald J. Williams David E. Rumelhart. Learning representations by back-propagating errors. *Nature*, 323, 1986.

[140] C. von der Malsburg. Self-organization of orientation sensitive cells in the striate cortex. *Kybernetik*, 14(2):85–100, Dec 1973.

[141] E. Oja. A simplified neuron model as a principal component analyzer. *J Math Biol*, 15(3):267–273, 1982.

[142] E. L. Bienenstock, L. N. Cooper, and P. W. Munro. Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. *J Neurosci*, 2(1):32–48, Jan 1982.

[143] A. J. Bell and T. J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Comput*, 7(6):1129–1159, Nov 1995.

[144] J. Triesch. A gradient rule for the plasticity of a neuron's intrinsic excitability . In *Int. Conf. on Artificial Neural Networks*, 2005.

[145] J. M. Fellous and C. Linster. Computational models of neuromodulation. *Neural Comput*, 10(4):771–805, May 1998.

[146] Obermayer, Blasdel, and Schulten. Statistical-mechanical analysis of self-organization and pattern formation during the development of visual maps. *Phys Rev A*, 45(10):7568–7589, May 1992.

[147] Risto Miikkulainen, James A. Bednar, Yoonsuck Choe, and Joseph Sirosh. *Computational maps in the visual cortex.* Springer, 2005.

[148] Andrew P Davison, Daniel Brüderle, Jochen Eppler, Jens Kremkow, Eilif Muller, Dejan Pecevski, Laurent Perrinet, and Pierre Yger. Pynn: A common interface for neuronal network simulators. *Front Neuroinformatics*, 2:11, 2008.

[149] G. B. Ermentrout and J. D. Cowan. A mathematical theory of visual hallucination patterns. *Biol Cybern*, 34(3):137–150, Oct 1979.

[150] J.-D. Ermentrout, G.-B.; Cowan. Large scale spatially organized activity in neural nets. *SIAM-J.-Appl.-Math.*, 38:1–21, 1980.

[151] Peter Tass. Cortical pattern formation during visual hallucinations. *Journal of Biological Physics*, 21:177–210, 1995. 10.1007/BF00712345.

[152] Peter Tass. Oscillatory cortical activity during visual hallucinations. *Journal of Biological Physics*, 23:21–66, 1997. 10.1023/A:1004990707739.

[153] P. C. Bressloff. *Methods and Models in Neurophysics: Lecture Notes of the Les Houches Summer School 2003*, chapter Pattern formation in visual cortex. Elsevier Science, 2004.

[154] Tanya I Baker and Jack D Cowan. Spontaneous pattern formation and pinning in the primary visual cortex. *J Physiol Paris*, 103(1-2):52–68, 2009.

[155] S. Coombes. Waves, bumps, and patterns in neural field theories. *Biol Cybern*, 93(2):91–108, Aug 2005.

[156] S. Amari. Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics*, 27, 1977.

[157] G B Ermentrout and J B McLeod. Existence and uniqueness of travelling waves for a neural network. *Proceedings of the Royal Society of Edinburgh*, 123A:461478, 1993.

[158] S. Coombes and M. R. Owen. Bumps, breathers, and waves in a neural network with spike frequency adaptation. *Phys Rev Lett*, 94(14):148102, Apr 2005.

[159] Rodica Curtu and Bard Ermentrout. Pattern formation in a network of excitatory and inhibitory cells with adaptation. *SIAM J. Appl. Dyn. Syst.*, 3:191–231, 2004.

[160] S. E. Folias and P. C. Bressloff. Breathers in two-dimensional neural media. *Phys. Rev. Lett.*, 95(20):208107, Nov 2005.

[161] Matteo Carandini Ian Nauhaus, Laura Busse and Dario L Ringach. Stimulus contrast modulates functional connectivity in visual cortex. *Nature Neuroscience*, 12:70–76, 2008.

[162] D. Durstewitz and J. K. Seamans. Beyond bistability: biophysics and temporal dynamics of working memory. *Neuroscience*, 139(1):119–133, Apr 2006.

[163] Z. F. Kisvrday, E. Tth, M. Rausch, and U. T. Eysel. Orientation-specific relationship between populations of excitatory and inhibitory lateral connections in the visual cortex of the cat. *Cereb Cortex*, 7(7):605–618, 1997.

[164] Judith S. Law. *Modeling the Development of Organization for Orientation Preference in Primary Visual Cortex.* PhD thesis, School of Informatics, The University of Edinburgh, 2009.

[165] Kenichi Ohki, Sooyoung Chung, Prakash Kara, Mark Hbener, Tobias Bonhoeffer, and R. Clay Reid. Highly ordered arrangement of single neurons in orientation pinwheels. *Nature*, 442(7105):925–928, Aug 2006.

[166] Sepideh Sadaghiani, Guido Hesselmann, Karl J Friston, and Andreas Kleinschmidt. The relation of ongoing brain activity, evoked neural responses, and cognition. *Front Syst Neurosci*, 4:20, 2010.

[167] Junji Ito, Andrey R Nikolaev, and Cees van Leeuwen. Dynamics of spontaneous transitions between global brain states. *Hum Brain Mapp*, 28(9):904–913, Sep 2007.

[168] Rubén Moreno-Bote, John Rinzel, and Nava Rubin. Noise-induced alternations in an attractor network model of perceptual bistability. *J Neurophysiol*, 98(3):1125–1139, Sep 2007.